2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Longitudinal modeling of antibody dynamics during herpes zoster infection

Supervisor :
Prof. dr. Niel HENS

Supervisor :
Dr. BENSON OGUNJIMI

Mai Phuong Thao Tran

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Longitudinal modeling of antibody dynamics during herpes zoster infection

Supervisor :
Prof. dr. Niel HENS

Supervisor :
Dr. BENSON OGUNJIMI

Mai Phuong Thao Tran

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Hasselt University

Master Thesis

Longitudinal modeling of antibody dynamics during Herpes Zoster infection

Author:
Thao Mai Phuong Tran

Supervisor:
Prof. Dr. Niel Hens
Dr. Ogunjimi Benson

September 1, 2015

Acknowledgements

From the bottom of my heart, I would like to express my great gratitude to the Belgian Technical Cooperation (BTC) for providing me with a scholarship to study in Belgium. It was an amazing opportunity to broaden my knowledge in such a wonderful field of statistics, sharpen my skills and share my experiences with people from all over the world. My sincere thanks is given to Hasselt University, especially Censtat for helping me get acquainted with the new life, gain the most interesting but also challenging statistical knowledge ever. I further want to say million thanks to my supervisors, Prof.Dr. Niel HENS and Dr. Beson Ogunjimi for their continuous guidance, advice, encouragement and useful comments on my thesis. Also, I would be grateful to the Vietnamese community in UHasselt, especially my boyfriend and Vietnamese friends who deliberately support me not only in the daily life but also in studying process. Finally, my heart is devoted to my family in Vietnam. They are the reason why I never give up even life is sometimes difficult.

Contents

List of Tables

List of Figures

Terms and Abbreviations

HZ	Herpes Zoster
PY	Person-year
PHN	Postherpetic neuralgia
VZV	Varicella zoster virus
BM	Bone marrow
HAV	Hepatitis A Virus
FDCs	Follicular dendritic cells
TLR	Toll-like receptor
IWRES	Individual weighted residuals
NPDE	Normalized prediction distribution errors
PWRES	Population weighted residuals
GLMM	Generalized Linear Mixed Model
GEE	Generalized Estimating Equations
SAEM	Stochastic Approximation Expectation Maximization
EM	Expectation Maximization
CI	Confidence Interval

1 Introduction

Herpes zoster (HZ), known as shingles, is a skin condition that caused by the reactivation of a latent varicella zoster virus. This virus is known to cause chickenpox (Lee et al., 2013). After a person recovers from chickenpox, the virus stays inactive in the body and can reactivates years later for reasons not fully known, causing shingles (CDC, 2015). Every year, the number of new cases of HZ in the United States is approximately 1 million (Lee et al., 2013). In Europe, the overall annual HZ incidence varied from 2.0 - 4.6/1000 person-years (PY). The incidence was lower in Iceland, Germany and Switzerland (around 2/1000 PY), medium in the United Kingdom, the Netherlands and France (around 3/1000 PY) and higher in Belgium, Spain and Italy (around 4/1000 PY) (Pinchinat et al., 2013). One major complication with HZ is postherpetic neuralgia (PHN), which is dermatomal pain after the resolution of rash that can be exquisitely painful and last months to even years (Lee et al., 2013).
It is now a challenge to understand the mechanisms that connect to humoral immunity's long-term endurance. This knowledge is pivotal for further studies in the field of immunology and health policy(Andraud et al., 2012). Currently, there are not many studies which provides insight knowledge of the quantitative assessment underlying the biological kinetics of immunity activities in human body, especially for HZ disease. The current study provides some basic results which could be contributed to further researches on immunology, especially in the field of humoral and vaccine immunity, not only for HZ disease but also could be a beginning step to further study immunity for other diseases. The analysis uses unbalanced repeated measurements from a study enrolling 61 patients to investigate antibody dynamics during HZ infection. We use the imprinted lifespan model proposed by Amanna and Slifka (2010) which was then applied to analyze the dynamics of plasma cell and antibody populations illustrated for hepatitis A virus by Andraud et. al. (2012). This model employs an Ordinary Differential Equation (ODE) where it describes the change in subpopulations of short-lived plasma cells, long-lived plasma cells and antibodies. Our model assumes that the long-lived plasma cells have a relatively long lifespan hence, this subpopulation is considered constant (this is called asymptotic models). Later on, the effects of some important covariates, including AGE, AV, VL (stand for age of patients at entry, antiviral usage and viral load obtained at the first measurement for each patient) are incorporated into the model. Furthermore, the reduced asymptotic models (model with and without covariates) assuming the average decay rate of short-lived plasma cells much shorter compared to that of antibody are fitted. The results under these models are than compared to the asymptotic models. Non-linear mixed effects model involving ODE is fitted using Monolix software.
As mentioned earlier, PHN is a common complication of HZ. Along with fitting a mathematical kinetics model to antibody data, we additionally aim to investigate the longitudinal relationship between the probability of having PHN over time and antibody titers. Firstly, a GLMM model where $time, LOGAB$ ($LOGAB$ is the log_{10} scale of antibody titers) are covariates is fitted to answer the question of interest. It is documented that age is a factor which might affect the appearance of PHN (Dworkin and Schmader, 2001). Moreover, the usage of antiviral drug plays a role to the control of PHN (Dworkin

and Schmader, 2003). Based on these backgrounds, a model considering AGE, AV as additional covariates is fitted later. We use a random effect (partial) proportional odds model to deal with this problem as the response is an ordinal variable.

This Page Intentionally Left Blank

2 Literature review

2.1 Herpes Zoster virus diseases

Herpes zoster (HZ) is a distinctive syndrome caused by reactivation of varicella zoster virus (VZV). This reactivation occurs when immunity to VZV declines because of aging or immunosuppression. HZ can occur at any age but most commonly affects the elderly population (Sampathkumar et al., 2009b). The diagnosis of HZ is mainly made clinically, except in patients with atypical manifestations or certain complications, such as central nervous system involvement, in which laboratory virologic testing is required (Bader, 2013).

2.2 Postherpetic neuralgia (PHN)

Post-herpetic neuralgia (PHN) is defined as pain persisting more than 3 months after the rash has healed. PHN is a debilitating and difficult to manage consequence of HZ (Sampathkumar et al., 2009a). The risk of PHN after HZ increases with age (Sampathkumar et al., 2009a; Gerhson, 1996). More than 50% of HZ patients older than 60 years will develop PHN and this may persist for months, even years (Gerhson, 1996). PHN is often severe and, unfortunately, refractory to most forms of treatment. As a result, established PHN is usually difficult to manage, often leading to serious morbidity, depression and high costs in healthcare resources (Gerhson, 1996).

2.3 Mathematical models of antibody kinetics

Humoral immunity following vaccination or infection is mainly derived from two types of cells: memory B cells and plasma cells (Amanna and Slifka, 2010). Amanna and Slifka suggests 6 models of sustained humoral immunity. They are chronic infection/cross-reactivity, repeated infection/vaccination, persisting antigen, polyclonal stimulation, competition for bone marrow (BM) and imprinted lifespan. The first four models belong to memory B-cell (MBC)-dependent and -independent models. They have been developed to explain how long-term antibody responses are maintained (Amanna and Slifka, 2010).

Chronic infection/cross-reactivity model: "Under this model assumption, chronic infection or cross-reactivity to either self or environmental antigens is expected to stimulate memory B cells to proliferate and differentiate into antibody-secreting daughter cells and result in increasing antibody responses over time due to continuous stimulation and accumulation of memory B cells and plasma cells" (Amanna and Slifka, 2010).

Repeated infection/vaccination model:" The model assumes that repeated infection or booster vaccination will likely lead to periodic increases in antigen-specific memory B-cell activation and subsequent increases in antibody responses that would decline during the intervening periods between outbreaks or vaccinations" (Amanna and Slifka, 2010).

Persisting antigen model: "In this model, there is persisting antigen in the form of antibody: antigen immune complexes on the surface of follicular dendritic cells (FDCs) will stimulate memory B cells in an antigen-specific manner, resulting in antibody response that will decline at the rate of antigen decay or consumption by the memory B-cell pool" (Amanna and Slifka, 2010).

Polyclonal stimulation: "It is said that non-antigen-specific polyclonal memory B-cell stimulation by Toll-like receptor (TLR) engagement or bystander T-cell activation will trigger antibody responses to spike during heterologous infections or vaccinations and increase antibody response to all pre-existing antibody specificities" (Amanna and Slifka, 2010).

However, these models, according to Amanna and Slifka, are not relevant to capture the progression of antibody titers over time after exposure to viral or vaccine antigens (Amanna and Slifka, 2010; Andraud et al., 2012). Furthermore, long-term antibody responses may be maintained by long-lived plasma cells (PCs) (Amanna and Slifka, 2010) which is not reflected in these four models. Therefore, Amanna and Slifka (2010) propose two other models where plasma cells are recognized as an independent B-cell subpopulation. This subpopulation is long-lived even without the replenishment by memory B-cells (Amanna and Slifka, 2010; Andraud et al., 2012), they are the competition for bone marrow (BM) model and the imprinted lifespan model.

Competition for BM model: "This model is based on plasma cell competition for space in the bone marrow in which pre-existing plasma cells are dislodged by incoming plasmablasts, and antibody responses decline as a function of plasma cell displacement. Amanna and Slifka argued that because there was finite space in the bone marrow, this model would suggest that antibody responses will decline more rapidly during advanced age as a function of increased competition in the bone marrow compartment" (Amanna and Slifka, 2010).

Imprinted lifespan model: "The model is based on the theory that plasma cells are imprinted with a specified lifespan, which is determined during the induction phase of the antigen-specific antibody responses" (Amanna and Slifka, 2010).

2.4 Plasma-cell imprinted lifespan model - An application by Andraud et al.

The "plasma-cell imprinted lifespan model" proposed by Amanna and Slifka was used by Andraud et al. (2012) to estimate persisting endurance of anti-HAV antibodies from two 10-year follow-up studies conducted in adults who were vaccinated with inactivated hepatitis A vaccines. The basic idea in this model is that it accounts for the dynamics of plasma cell (P) and antibody (A) populations. The plasma cell population is divided into two subpopulations, which are short- and long-lived cells based on their lifespans, denoted as P_s and P_l. Under this model, these plasma cells are assumed to have no renewal, meaning that they decline over time. Additionally, lifespan of antibody is assumed to be relatively short compared to that of plasma cell. The model is said to demonstrate the kinetics of plasma cell populations (Andraud et al., 2012).

The complete model

Following these presumptions, the dynamics of plasma cells and antibody are described by the Ordinary Differential Equations (ODE) below:

$$\frac{dPs}{dt} = -\mu_s Ps \tag{1}$$

$$\frac{dPl}{dt} = -\mu_l Pl \tag{2}$$

$$\frac{dA}{dt} = \varphi_s Ps + \varphi_l Pl + \mu_A A \tag{3}$$

This is called a complete model, where μ_s, μ_l, μ_A are the average decay rates of P_s, P_l and A; φ_s, φ_l are the average production rates of A by short- and long-lived plasma cells. Let denote P_s^0, P_l^0, A_0 as the initial population sizes of P_s, P_l and A, the ODE system has the analytical solution:

$$\begin{cases} P_s(t) = P_s^0 e^{-\mu_s t} \\ P_l(t) = P_l^0 e^{-\mu_l t}, \\ A(t) = \dfrac{\phi_s}{\mu_A - \mu_s} e^{-\mu_s t} + \dfrac{\phi_l}{\mu_A - \mu_l} e^{-\mu_l t} + (A_0 - \dfrac{\phi_s}{\mu_A - \mu_s} - \dfrac{\phi_l}{\mu_A - \mu_l}) e^{-\mu_A t} \end{cases}$$

where $\phi_s = \varphi_s P_s^0$ and $\phi_l = \varphi_l P_l^0$.

The asymptotic model

Assuming that the lifespan of short-lived plasma cells is infinity ($\mu_l = 0$), Andraud et. al (2012) proposed a model where the asymptotic total antibody production rate is a constant different from zero ($\frac{\phi_l}{\mu_A}$). The analytical solution for antibody at time point t becomes:

$$A(t) = \frac{\phi_l}{\mu_A} + \frac{\phi_s}{\mu_A - \mu_s} e^{-\mu_s t} + (A_0 - \frac{\phi_l}{\mu_A} - \frac{\phi_s}{\mu_A - \mu_s}) e^{-\mu_A t}$$

The plasma cell driven kinetic model

This model assumes that the antibody lifespan is relatively short compared to those of short- and long-lived plasma cells ($\mu_s, \mu_l << \mu_A$). "The antibody kinetics can be considered as being an immediate reflection of the underlying kinetics of plasma cell populations" (Amanna and Slifka, 2010; Andraud et al., 2012). The analytical solution for antibody is then given by:

$$A(t) = \beta_s e^{\mu_s t} + \beta_l e^{\mu_l t}$$

where $\beta_s = \dfrac{\phi_s}{\mu_A}$, $\beta_l = \dfrac{\phi_l}{\mu_A}$.

3 Data

The original data was collected on 61 patients being diagnosed as HZ infection in general practitioners. The data are repeated and unbalanced in the sense that the number and time points of measurements vary from subjects to subjects. There are total 233 observations and 9 variables. Table 1 gives the details of all variables collected in the data set.

Table 1: Variables explanation

Variable	**Type**	**Explanation**
ID	NA	Unique Identification of patients
AGE	Continuous	Age of patients at time point 0
GEN	Binary	Genetic characteristics of patients (1 or 0)
AV	Binary	Indication of using antiviral treatment (1 if Yes and 0 if No)
DUR	Continuous	Duration of using antiviral drug
TIME	Time	Time point of measurement
AB	Continuous	Antibody level at each measurement
CMV	Binary	Indication of Cytomegalovirus (1 or 0)
VL	Continuous	Viral load at each measurement
PHN	Categorical	Postherpetic neuralgia status at each visit (0 if no pain, 1 if dyskinesia but no pain and 3 if PHN)

4 Methodology

4.1 Data Exploratory

Data exploratory is applied to get general information about the data. The individual profile and mean structure are explored using the methods applied to unbalanced longitudinal data.

4.2 Mathematical models of antibody kinetics - Application into the current data set

4.2.1 Why the approach of Andraud et al. might not reasonable for the current data set

The mathematical model proposed by Andraud et al. assumes a simple decay trend over time of antibody levels in vaccinated adults with inactivated hepatitis A vaccines. This assumption seems to be unfeasible in our study. To make clear this argument, it is necessary to have a short overview about the correlation of antibody titers with various phases of vaccine or virus response (humoral immunity). The initial antigen exposure (virus, bacteria,...) elicits an extrafollicular response resulting in the rapid appearance of low IgG antibody titers. When B cells proliferate in germinal centers and differentiate into plasma cells, IgG antibody titers increase up to a peak value usually reached 4 weeks after immunization. Due to the short lifespan of these plasma cells, it results in a rapid decline of antibody titers, which might eventually return to baseline levels. In the secondary immune responses, booster exposure to antigen reactivates immune memory and results in a rapid (less than one week) increase of IgG antibody titer. Short-lived plasma cells maintain peak antibody levels during a few weeks after which serum antibody titers decline initially with the same rapid kinetics as following primary immunization. Long-lived plasma cells having reached survival niches in the bone marrow continue to produce antigen-specific antibodies, which then decrease with slower kinetics. It is noted that this generic pattern may not apply to live vaccines triggering longterm IgG antibodies for extended periods of time (Siegrist, 2015).
In the paper of Andraud et al., blood samples were taken before vaccination, between primary and after booster administration. As the purpose of their study is to investigate the durable endurance of antibodies after a full vaccination program, they only limited the use of data at those time-points after boosting, *i.e.* at 1, 12, 18, 24, 30, 36, 42, 48, 50, 66, 78, 90, 102, 114 and 126 months after boosting. This might explain why a simple decay model is appropriate in this analysis.
Our study collected data from the time of diagnose of HZ for all 61 patients. As a result, a model that reflects a continuous exposure rather than a simple decay model could be more appropriate since it is assumed that there is continuous exposure in the population.

4.2.2 A proposed model without covariates

Having realized that the model proposed by Andraud et al. is appropriate for modeling a simple decay problem, we adapt their approach so that it can be applied to model the antibody kinetics when there is continuous exposure. Our model is developed based on the idea of Andraud et al., but taking into account the fact that antibody levels could be firstly increased to a peak (before a time point t_1), and then decline over time (after time point t_1).

This proposed model is reasonable with the biological ground. As we have known, there are primary and secondary responses to antigen exposure. The initial immune response is called the primary response. When an antigen appears again, it triggered a more extensive and prolonged secondary response. During the primary response, the antibody titer (level of antibody activity) in the plasma does not peak until 1 or 2 weeks after the initial exposure. If the individual is no longer exposed to the antigen, the antibody concentration decreases. In the secondary response which is characterized by a very rapid increase in IgG antibody concentration and titer, the antibody titer rises to a much higher levels than those of primary response. Antibody activity remains elevated for an extended period after the second exposure to the antigen (Martini and Nath, 2011).

Based on these argument, we propose the model with two ODE systems:

Stage 1: Before time point t_1, antibody titer increases over time. We have the following ODE system which is a consideration between a mathematical/biological model and a statistical model, under the assumption that AB decay only starts from a time point t_1.

$$\frac{dPs}{dt} = -\mu_s Ps \tag{4}$$

$$\frac{dPl}{dt} = -\mu_l Pl \tag{5}$$

$$\frac{dA}{dt} = \varphi_s Ps + \varphi_l Pl \tag{6}$$

Stage 2: From time point t_1, antibody titer starts declining over time. The following ODE system is employed:

$$\frac{dPs}{dt} = -\mu_s Ps \tag{7}$$

$$\frac{dPl}{dt} = -\mu_l Pl \tag{8}$$

$$\frac{dA}{dt} = \varphi_s Ps + \varphi_l Pl - \mu_A A \tag{9}$$

where μ_s, μ_l and μ_A respectively stand for the average decay rates of short-, long-lived plasma cells antibodies in the body. φ_s, φ_l are the production rates of antibodies by short- and long-lived plasma cells. Ps^0, Pl^0, A_0 are subsequently the number of short-lived plasma cells, long-lived plasma cells and antibody titers at time point $t = 0$.

By fitting this model in Monolix, it turns out that the parameter t_1, μ_l, φ_l are very close

to 0. This motivates our choice to fit a model with only one ODE system, and assume that $\mu_l = 0$, i.e, the average decay rate of long-lived plasma cells could be forever ignored.

i. The Asymptotic Model

The final ODE system to be considered in this case is given as following:

$$\frac{dPs}{dt} = -\mu_s Ps \tag{10}$$

$$\frac{dA}{dt} = \varphi_s Ps + \phi_l - \mu_A A \tag{11}$$

This model is called the asymptotic model which is one of three models proposed by Andraud et al. (2012). This means that although the complete model proposed by Andraud et al. seems not be appropriate, the asymptotic one might show a better behavior. This model omits the dynamic equation $\frac{dPl}{dt} = -\mu_l Pl$, reflecting the fact that long-lived plasma cells are assumed to survive for a long period. It is said that the number of long-lived plasma cells that disappear could be ignored. Moreover, looking at the formula 11, it can be seen that now the parameter φ_l and Pl^0 are absorbed into one parameter ϕ_l reflecting the number of antibodies secreted by long-lived plasma cells at the baseline.
The ODE system has the analytical solution:

The asymptotic model:

$$A(t) = \frac{\phi_l}{\mu_A} + \frac{\phi_s}{\mu_A - \mu_s} e^{-\mu_s t} + (A_0 - \frac{\phi_s}{\mu_A - \mu_s} - \frac{\phi_l}{\mu_A}) e^{-\mu_A t} \tag{12}$$

where $\phi_s = \varphi_s Ps^0$.

ii. The reduced model

Assuming that $\mu_s << \mu_A$, i.e., the lifespan of short-lived plasma cells is relatively larger compared to that of antibody. The analytical solution 12 becomes:

The reduced asymptotic model:

$$A(t) = \frac{\phi_l}{\mu_A} + \frac{\phi_s}{\mu_A} e^{-\mu_s t} + (A_0 - \frac{\phi_s}{\mu_A} - \frac{\phi_l}{\mu_A}) e^{-\mu_A t} \tag{13}$$

Parameter estimation

Non-linear mixed effects model approach was employed to estimate population parameters $(\mu_s, \varphi_s, \mu_A\ ,\ Ps_0, A_0, \phi_l)$. The individual parameters (ψ_i) are assumed to have

log-normal distributions. That is $log(\psi_i) = log(\theta) + \eta_i$ where the subscript i represents the individual i, $log(\theta)$ is the fixed effect parameter representing the mean value of the parameter in the population and η_i is the random effect accounting for the inter-individual variability. η_i is assumed to have a normal distribution, that is $\eta_i \sim N(0, \omega^2)$. Individual parameter estimates are used to predict the individual antibody titer at each time point ($A_{pred,ij}$). The measured antibody titers are log10-transformed for the computational convenience since the range of this value is quite large. The model for log10-transformed data is described as following:

$$log_{10}(A_{Obs,ij}) = log_{10}(A_{Pred,ij}) + \epsilon_{ij} \tag{14}$$

The constant error model is used. ϵ_{ij} is assumed to be normally distributed, that is $\epsilon_{ij} \sim N(0, a^2)$. While ω^2 reflects inter-individual variability, a^2 quantifies the residual variability. Population parameters are estimated using SAEM algorithm in Monolix (Lixoft, 2014).

Model diagnostic

We use AIC for the purpose of selecting the most appropriate model. Goodness of fit of models were checked by using plots for individual predictions, the individual weighted residuals (IWRES) and normalized prediction distribution errors (NPDE). This makes sense since population based diagnostics were not very instructive (Andraud et al., 2012; Karlsson and Savic, 2007).

4.2.3 A proposed model with covariates

i. The Asymptotic Model

This model takes into account the covariates which are viral load (VL), AGE of patients, antiviral drug used (AV), duration of using antiviral drug (DUR), genetic characteristics (GEN) and CMV. Since AV and DUR give more or less the same information, i.e., if the antiviral drug is used then the usage duration is larger then 0, otherwise, the usage duration is 0, it is decided here to use only one variable AV instead of using both AV and DUR variables. The covariates are assumed to influence the parameters μ_A and μ_s. The model for individual parameters $\mu_{A,i}$ and $\mu_{s,i}$ are described as following (More motivation could be found under Section 5.2.2):

$$\begin{aligned}
log(\mu_{s,i}) &= log(\mu_{s,pop}) + \beta_{\mu_s,AGE} AGE_i + \beta_{\mu_s,AV} AV_i + \\
&\quad \beta_{\mu_s,VL}(log10(VL_i) - log10(\overline{VL_i})) + + \eta_{\mu_s,i} \\
\eta_{\mu_s,i} &\sim N(0, \omega^2_{\mu_s}) \\
log(\mu_{A,i}) &= log(\mu_{A,pop}) + \beta_{\mu_A,AGE} AGE_i + \beta_{\mu_A,AV} AV_i + \\
&\quad \beta_{\mu_A,VL}(log10(VL_i) - log10(\overline{VL_i})) + \eta_{A_0,i} \\
\eta_{A_0,i} &\sim N(0, \omega^2_{\mu_A})
\end{aligned}$$

Since VL is measured at every visits (i.e changes within occasion), only the first valid value is taken by Monolix. In the data set, there are two missing values of viral load for patient ID 2 at visit 3 and patient ID 34 at visit 2. So viral load values are taken into analysis by using the measurements at the first visit, i.e., at time point $t = 0$ for all subjects.

ii. The reduced asymptotic model

In the reduced asymptotic model, we assume that three covariates AGE, AV and VL influence to the individual parameter estimates of μ_s, μ_A and ϕ_s. The motivation for this choice could be found under Section 5.2.2. The models for individual parameters $\mu_{A,i}, \mu_{s,i}, \phi_{s,i}$ are specified as following:

$$
\begin{aligned}
log(\mu_{s,i}) &= log(\mu_{s,pop}) + \beta_{\mu_s,AV} AV_i + \beta_{\mu_s,AGE} AGE_i + \beta_{\mu_s,VL}(log_{10}(VL_i) - log_{10}(\overline{VL})) + \eta_{\mu_s,i} \\
\eta_{\mu_s,i} &\sim N(0, \omega^2_{\mu_s}) \\
log(\mu_{A,i}) &= log(\mu_{A,pop}) + \beta_{\mu_A,AV} AV_i + \beta_{\mu_A,AGE} AGE_i + \beta_{\mu_A,VL}(log_{10}(VL_i) - log_{10}(\overline{VL_i})) + \eta_{A_0,i} \\
\eta_{A_0,i} &\sim N(0, \omega^2_{\mu_A}) \\
log(\phi_{s,i}) &= log(\phi_{s,pop}) + \beta_{\phi_s,AV} AV_i + \beta_{\mu_s,AGE} AGE_i + \beta_{\mu_s,VL}(log_{10}(VL_i) - log_{10}(\overline{VL_i})) + \eta_{\phi_s,i} \\
\eta_{\phi_s,i} &\sim N(0, \omega^2_{\phi_s})
\end{aligned}
$$

4.3 Model to investigate the relationship between PHN and antibody titers

At visit 2, 3 and 4, patients were evaluated by physicians to see if they had experienced PHN. The clinical judgement at visit 1 was not taken, but it was assumed that all patients had had no pain at this visit, i.e, $PHN = 0$. To investigate the relationship between antibody titers and having PHN, a model takes into account the inter-individual variability between measurements is considered. Having realized the fact that the response variable PHN is an ordinal one, a proportional odds ratio model could be appropriate. Both marginal and mixed effects proportional odds ratio model could be employed. Another proposed model for ordinal responses is the continuation-ratio logits model. However, this model might be convenient and useful only for subjects that gradually go through a number of states, where no return is possible (for example, cancer stages) (Molenberghs and Verbeke, 2005), or when a sequential mechanism, for example, survival through various age periods, determines the response (Agresti, 2002). That is why in this problem, we only consider to use proportional odds model.

4.3.1 Proportional Odds Model assumption

The proportional odds model is of the form (more motivation could be found under Section 5.3.1):

$$logit[P(PHN_{ij} \leq k)] = \beta_0 + \beta_1 t_{ij} + \beta_2 LOGAB_{ij} + \beta_3 t_{ij} LOGAB_{ij} + \beta_4 t_{ij}^2$$
$$k = 0, 1, 2.$$

If the model deviates from the proportional odds assumption, they can take on the form of either a partial proportional odds model (a model where a subset of covariates is assumed to have different parameters for each logit) or a non-proportional odss model (a model where all covariates are assumed to have different parameters for each logit) (Hedeker and Gibbons, 2006). We use the likelihood ratio test to see which models fit the data best.

4.3.2 Random-Effects Model with proportional odds assumption

Model with LOGAB as covariate
A generalized linear mixed model (GLMM) which is the most frequently used random-effects model in the context of discrete repeated measurements (Molenberghs and Verbeke, 2005) is considered. First of all, a model with random intercept and random slope is fitted but the convergence could not be achieved. We then consider the random intercept only model. The proportional odds model takes the form:

$$logit[P(PHN_{ij} \leq k|b_i)] = \beta_0 + b_i + \beta_1 t_{ij} + \beta_2 LOGAB_{ij} + \beta_3 t_{ij} LOGAB_{ij} + \beta_4 t_{ij}^2 \quad (15)$$
$$k = 0, 1, 2.$$

The model assumes that the logit evolves linearly over time. Here, b_i is the random intercept reflecting the variation of the logit of subject i from the population. It is assumed that $b_i \sim N(0, \sigma_b^2)$.

Model with additional covariates: AGE, AV
As mentioned earlier, some recent researches have showed that older age, greater acute pain during HZ and greated rash severity are identified as risk factors for PHN (Whitley et al., 1999; Dworkin and Schmader, 2001, 2003). Moreover, it is concluded that early diagnosis with HZ and treatment with antiviral drugs decreases the risk of PHN (Sampathkumar et al., 2009a). In this analysis, the baseline effects of age and antiviral treatment indication are considered as additional baseline covariates in the model. It takes of the form:

$$logit[P(PHN_{ij} \leq k|b_i)] = \beta_0 + b_i + \beta_1 t_{ij} + \beta_2 LOGAB_{ij} + \beta_3 t_{ij} LOGAB_{ij} + \beta_4 t_{ij}^2 +$$
$$\beta_5 AGE_{ij} + \beta_6 AV_{ij} + \beta_7 AGE_{ij} t_{ij} + \beta_8 AV_{ij} t_{ij} \quad (16)$$
$$k = 0, 1, 2.$$

Again, b_i is the random intercept representing the within-subject variability, assumed to have normal distribution $b_i \sim N(0, \sigma_b^2)$.

4.4 Software

The data analysis is proceeded using Monolix software version 4.3 and SAS version 9.3. Monolix uses the SAEM algorithm to estimate the maximum likelihood estimator for population parameters. SAEM is a stochastic extension version of Expectation-Maximization algorithm (Lavielle, 2015). In non-linear mixed effects model, the regression function f does not linearly depend on the random effects (individual parameters), the E step in the Expectation-Maximization (EM) cannot be performed in a closed-form. The SAEM algorithm replaces the E step in EM algorithm by a stochastic procedure including simulation step and stochastic approximation. Convergence of SAEM can strongly depend on the initial guess if the likelihood possesses many local maxima. The simulated annealing version of SAEM improves the convergence of the algorithm toward the global maximum of likelihood (Lixoft, 2014).
A significance level of 5% is used for decision making.

5 Results

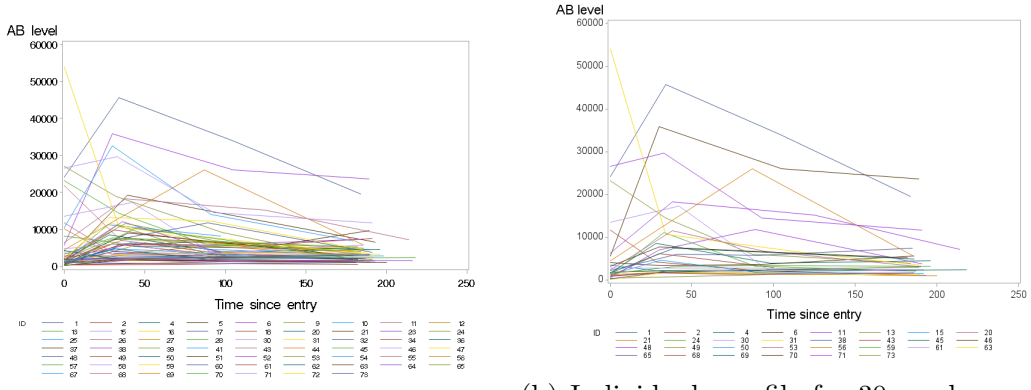
5.1 Data Exploratory

Our analysis uses data from a longitudinally unbalanced study. The time point $t = 0$ indicating the baseline when subjects were diagnosed as HZ infection at general practitioners. After successfully obtaining informed consent form to participate into the study, these patients were followed up with a longest period of 240 days. The maximum number of measurements per each subject is 4 and there is only one patient (ID = 55) has one measurement. There are four missing measurements for AB and two missing measurements for VL. Table 2 shows summary statistics for AB and VL. It can be seen that the values of VL vary a lot from measurements to measurements with very high standard deviation.

Table 2: Summary statistics for AB and VL

Var	N	Mean	SD	Min	Max
AB	229	6116.27	7751.95	121	53952
VL	231	5440.82	15434.93	0	151685

5.1.1 Individual profile



(a) Individuals profile for all subjects

(b) Individuals profile for 30 random subjects

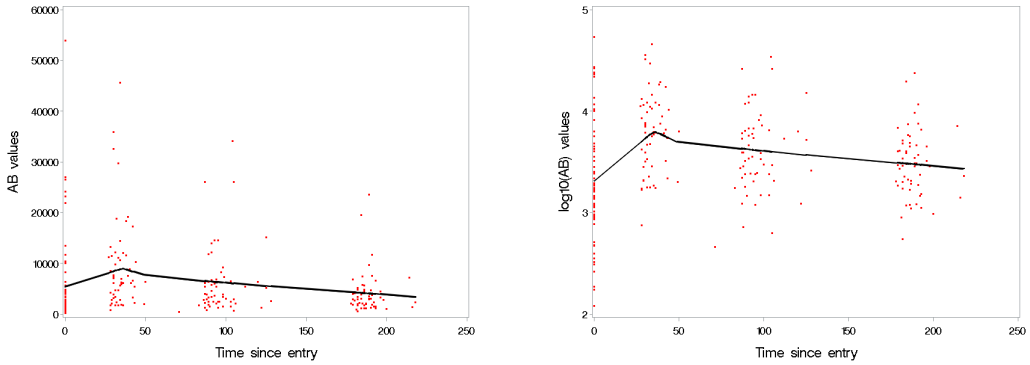
Figure 1: Individuals profile: The left panel shows the profile for all 61 subjects; the right panel shows the one for 30 randomly selected subjects.

Figure 1 shows the individuals profile for all subjects (the left panel) and for 30 randomly selected subjects (the right panel). It is seen that most subjects have low starting antibody levels ($t = 0$), increasing at the next time point and decreasing later on. There are some subjects that have very high antibody levels at the beginning ($t = 0$),

then decreasing over time. These subjects might come to the general practitioners very late after having the symptoms. Hence, the antibody levels at time point 0 are very high.

5.1.2 Mean structure

The mean structure of antibody levels as well as log_{10} of antibody levels over time are shown in the Figure 2. The left panel shows the plot for antibody levels and the right panel shows the plot for log_{10} of antibody levels. It can be seen that antibody levels (or log_{10} of antibody titers) firstly increase from time point $t = 0$ to the pick at around time point $t = 40$ and then decrease onward.



(a) Mean structure of antibody levels over time

(b) Mean structure of log_{10} of antibody levels over time

Figure 2: Mean structure: The left panel shows the mean structure of antibody levels over time; the right panel shows the mean structure of log_{10} of antibody levels over time.

5.2 Mathematical models of antibody kinetics

5.2.1 Model without covariates

i. The Asymptotic Model

It turns out that μ_A, ϕ_l are highly correlated. The correlation between the two parameters when running the model in Monolix is 0.95. To be able to account for this nearly perfect correlation, we let $\phi_l = b * \mu_A$. The updated analytical solution has the form:

The updated analytical solution (asymptotic model):

$$A(t) = b + \frac{\phi_s}{\mu_A - \mu_s} e^{-\mu_s t} + (A_0 - \frac{\phi_s}{\mu_A - \mu_s} - b) e^{-\mu_A t} \tag{17}$$

Now it is of interest to estimate P_s^0, A_0, μ_s, μ_A and b.

Model diagnostics

We run the model long enough (number of iterations per stage 1 and stage 2 in SAEM algorithm are subsequently 18000 and 4000) so as to reach convergence. The convergence could be assessed quickly by looking at Figure 10 (saem convergence plot) in the Appendix.

Figure 3 shows the plot between individual observations vs. model prediction, both are presented on log_{10} scale. It can be seen that the model shows a nice prediction ability as the points resemble the 45^0 line. Residual plots over time are shown in Figure 4, the right panel shows the normalized prediction distribution errors (NPDE), the left panel shows the individual weighted residual errors (IWRES). NPDE are a non-parametric version based on rank statistic of population weighted residuals (PWRES) which are defined as normalized difference between observations and their means. IWRES are estimates of standardized residuals based on individual predictions (Lavielle, 2015). The residuals plots confirm the suitability of the constant error model except there is a small concern when looking at the plot of IWRES over time as the residuals at time point 0 are quite small compared to the residuals at other time points. The plot between NDPE over time does not confirm this observation. The model gives an AIC of 113.05.

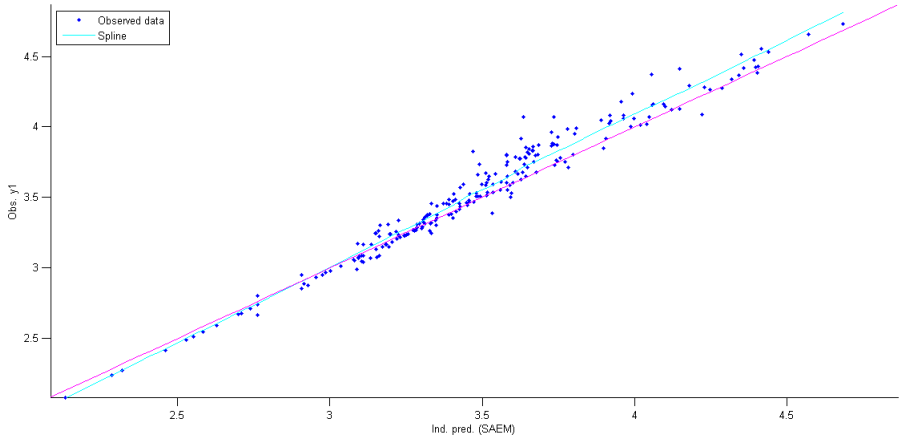


Figure 3: Individual observations v.s model prediction, both are presented on log_{10} scale (Asymptotic model without covariates).

Parameter estimation

The population parameter estimates and their standard errors are provided in Table 3. The standard errors (s.e (lin)) are estimated by using linearization method in Monolix. The estimated relative standard errors (r.s.e) are calculated by dividing the estimated standard errors by their corresponding values of the estimated parameters. The values of r.s.e are relatively small meaning that the model and data allow us to estimate the parameters well (Lavielle, 2015). The initial antibody titer is estimated around 2040 (mIU/ml). The estimated decay parameter for short-lived plasma cells (μ_s) is 0.0347

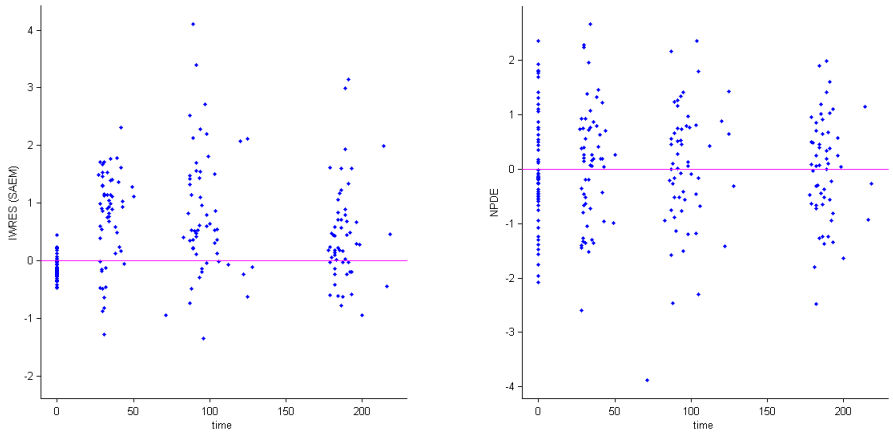


Figure 4: The residual plots: Individual weighted residuals (IWRES) v.s time (left) and normalized prediction distribution errors (NPDE) v.s time (right) (Asymptotic model without covariates).

and the decay parameter for antibody (μ_A) is higher (0.1180). Antibody decreases over time with a higher rate (nearly 3.4 times faster). Consequently, the estimated lifespan of short-lived plasma cells ($1/\mu_s$) is approximately 29 days which is longer compared to 8.5 days of estimated lifespan of antibody ($1/\mu_A$). It should be kept in mind that the model assumes the decay parameter of long-lived plasma cells is 0, hence this subpopulation could be considered constant over the investigated time.

ii. The Reduced Asymptotic Model

Let denote $\dfrac{\phi_l}{\mu_A} = b$. The updated analytical solution is given by:

The updated analytical solution (reduced asymptotic model):

$$A(t) = b + \frac{\phi_s}{\mu_A} e^{-\mu_s t} + (A_0 - \frac{\phi_s}{\mu_A} - b) e^{-\mu_A t} \tag{18}$$

Model diagnostics

The model is run in Monolix (number of iterations $K1 = 18000, K2 = 4000$) and the convergence is quickly assessed by looking at the SAEM convergence plot (the plot is not shown). The appropriateness of the model could be checked by looking at Figure 5 and Figure 12 in the Appendix. It can be seen that the model fits to the individual observations very well. The model gives AIC of 119.48.

Parameter estimation

Table 4 shows the population parameter estimates with their standard errors. The relative standard errors (r.s.e) are relatively small which indicates that the data and model

Table 3: Parameter estimates and their standard errors (Asymptotic model without covariate)

	Parm	**s.e. (lin)**	**r.s.e. (%)**
ϕ_s	763	120	15
μ_A	0.1180	0.0470	40
μ_s	0.0347	0.0110	33
b	2250	200	9
A_0	2040	360	18
ω_{ϕ_s}	0.4000	0.1700	42
ω_{μ_A}	1.6000	0.2400	15
ω_{μ_s}	0.7530	0.4000	53
ω_b	0.5770	0.0680	12
ω_{A_0}	1.3500	0.1300	10
a	0.1060	0.0120	11

Table 4: Parameter estimates and their standard errors (Reduced asymptotic model without covariate)

	Parm	**s.e. (lin)**	**r.s.e. (%)**
ϕ_s	1440	960	67
μ_A	0.3350	0.2400	72
μ_s	0.0137	0.0023	17
b	2030	240	12
A_0	2040	360	18
ω_{ϕ_s}	0.9680	0.5000	52
ω_{μ_A}	1.3100	0.4400	33
ω_{μ_s}	0.5530	0.1500	26
ω_b	0.6140	0.0840	14
ω_{A_0}	1.3500	0.1300	10
a	0.1090	0.0110	10

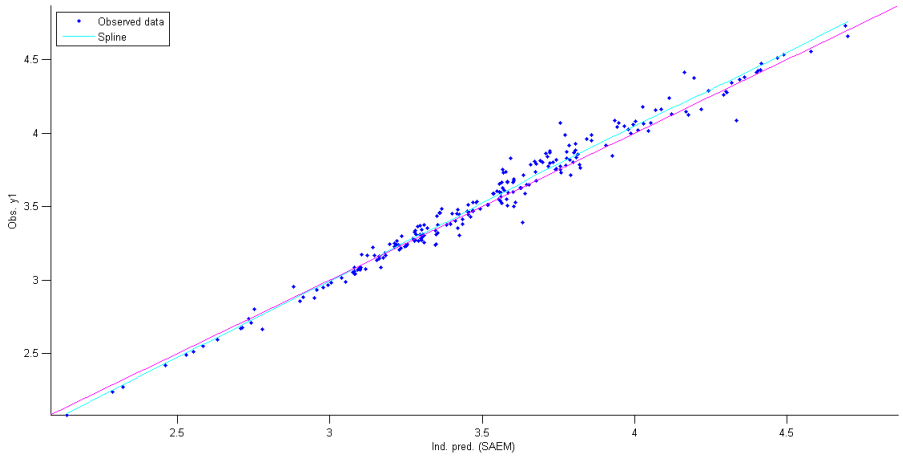


Figure 5: Individual observations v.s model prediction, both are presented on log_{10} scale (Reduced asymptotic model without covariates).

allow to estimate the parameters well. Comparing the outputs from Table 4 and Table 3, we see that the estimates of A_0 and b are quite close to each other. In the reduced asymptotic model, the estimate of ϕ_s is relatively higher compared to one in asymptotic model. The reduced asymptotic model gives estimate of μ_A higher than μ_s about 24 times. It is difficult to judge if the assumption of $\mu_s << \mu_A$ appropriate based on this observation. The lifespans of short-lived plasma cells and antibodies are respectively about 73 days and 3 days.

5.2.2 Model with Covariates

i. The Asymptotic Model
Model diagnostics
The model with all covariates as specified in Section 4.2.3 is fitted in Monolix with a large number of iterations. This model does not converge in Monolix may be due to overparameterization. It is then of interest to take into account three most important covariates including AGE, VL, AV. First of all, we fit the model where we assume these three covariates connecting to ϕ_s, μ_s, μ_A and A_0 (We call it Model (i)). This model gives AIC of 95.35 and $-2LogL = 49.35$. Secondly, models where covariates are assumed to influence a subset of any three parameters are fitted. All models give AICs larger than the one obtained from Model (i). Hence, Model (i) is more appropriate to our data at this stage. Thirdly, we fit the model where we assume these three covariates connecting to a subset of any two parameters. Only the model where it is assumed that these covariates influence individual estimates of μ_s, μ_A gives lower AIC compared to 95.35. This model is considered for making inference.
The convergence of this model could be quickly checked by looking at Figure 11 in the Appendix. Moreover, Figure 6 shows a good individual prediction ability since all the

points lie around the 45^0 line. Compared to the model which does not take into account the covariates, this model appears to perform better prediction ability.

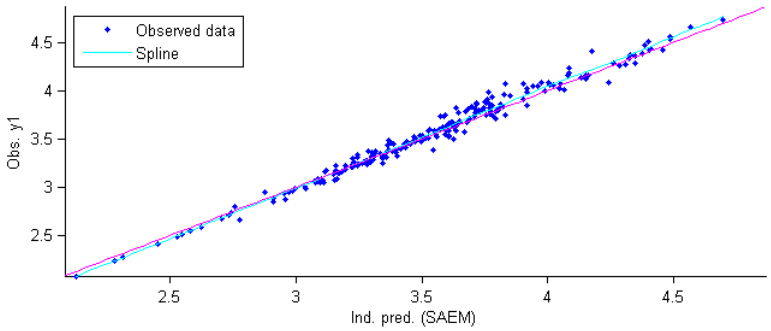


Figure 6: Individual observations v.s model prediction, both are presented on $log10$ scale (Asymptotic model with three covariates).

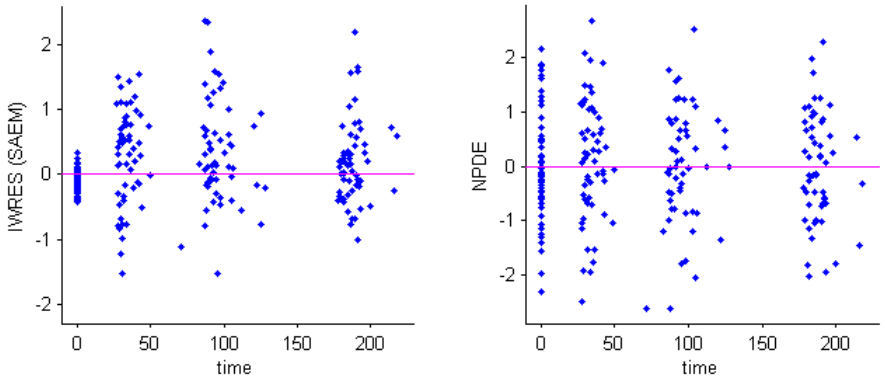


Figure 7: The residual plots: Individual weighted residuals - IWRES v.s time (left) and normalized prediction distribution errors - NDPE v.s time (right) (Asymptotic model with three covariates).

Figure 7 shows the residual plots. While the IWRES are quite small at time point 0, it shows no special pattern at the later time (left panel). The plot of NDPE over time seems to be quite reasonable under the constant error model assumption. Hence, the assumption of the model could be relaxed and the parameter estimation could be obtained based on this model. The model gives an AIC of 94.87.
Performing a global test by using the means of likelihood ratio test, we have $p - value < 0.0001$ (the $-2LogL$ under the model without covariates is 91.05 and the $-2LogL$ under the model with three covariates is 60.87, number of degree of freedom is 6). It leads to the conclusion that the model with covariates are preferable. This could be illustrated by better prediction ability when comparing Figure 3 and Figure 6.
We also fit the model with only two covariates (AGE and VL, VL and AV, AV and

AGE) and performing the likelihood ratio test. It leads to the conclusion that the model with 3 covariates is more appropriate ($p - values$ are respectively $0.0129, 0.0013$ and 0.0003).

Parameter Estimation

The parameter estimates are provided in Table 5. The relative standard errors are reasonably small meaning that the model and data allow us to estimate parameters well. While p-values for $\beta_{\mu_s,AGE}, \beta_{\mu_s,AV}, \beta_{\mu_A,VL}$ and $\beta_{\mu_A,AV}$ are highly significant (with corresponding $p - values$ are $0.0044, < 0.0001, 0.0088$ and 0.0270), the $p - values$ for $\beta_{\mu_A,AGE}$ and $\beta_{\mu_s,VL}$ are not significant (0.90 and 0.84 respectively).

Table 5: Parameter estimates and their standard errors (Asymptotic model with three covariates AGE, VL, AV)

	Parm	**s.e. (lin)**	**r.s.e. (%)**	**p-value**
ϕ_s	765	190	25	-
μ_A	0.0480	0.0310	65	-
μ_s	0.0055	0.0023	42	-
b	2110	230	11	-
A_0	2030	360	18	-
ω_{ϕ_s}	0.6700	0.2100	31	-
ω_{μ_A}	1.0900	0.1800	17	-
ω_{μ_s}	0.2540	0.3200	126	-
ω_b	0.6220	0.0760	12	-
ω_{A_0}	1.3500	0.1300	9	-
a	0.1010	0.0099	10	-
$\beta_{\mu_A,AGE}$	0.0014	0.0110	811	0.9
$\beta_{\mu_A,AV1}$	1.0900	0.4900	45	**0.0270**
$\beta_{\mu_A,tVL}$	-0.4110	0.1600	38	**0.0088**
$\beta_{\mu_s,AGE}$	0.0362	0.0082	49	$<$ **0.0001**
$\beta_{\mu_s,AV1}$	-1.0700	0.3800	35	**0.0044**
$\beta_{\mu_s,tVL}$	-0.0244	0.1200	487	0.84

Comparing the estimated population parameters from Table 3 (asymptotic model without covariates) and Table 5 (asymptotic model with covarites), we can see that while the estimates of ϕ_s, A_0 and b seem stable, the estimates of μ_s, μ_A in model with covariates is much smaller compared to the one obtained in model without covariates. As a result, the lifespan of short-lived plasma cells and antibody in the model with covariates are longer, respectively, 182 and 21 days. The lifespan of short-lived plasma cells is much longer compared to that of antibody in two models.

From the estimates of covariates, it is seen that the transformed viral load influences negatively to the estimates of the $\mu_{A,i}$ (with $\beta_{\mu_A,tVL} = -0.4110$) while age has a positive effect to the estimates of $\mu_{s,i}$. The usage of antiviral drug ($AV = 1$) affects positively to the estimate of μ_A ($\beta_{\mu_A,AV1} = 1.0900$) but negatively to the estimate of μ_s ($\beta_{\mu_s,AV1} = -1.0700$) indicating that these subjects have shorter lifespan of antibody but longer lifespan of short-lived plasma cells.

ii. The Reduced Asymptotic Model

Model diagnostics

For the reduced asymptotic model, we still consider three most important covariates, namely AGE, VL and AV. We fit the model where we assume these covariates affecting four parameters $(\phi_s, \mu_A, \mu_s, A_0)$. The converged model in Monolix gives an AIC of 115.06 and $-2LogL = 69.06$ (Model a). Next, we fit the model where three covariates are assumed to affect the estimates of subset of any three parameters. The model where the covariates assumed to influence the estimates of μ_A, μ_s and ϕ_s (Model b) gives smallest AIC (97.98) which is even smaller compared to the AIC obtained from Model a. Furthermore, we fit the models assuming that three covariates only affect the estimates of subset of any two parameters. These models give AICs higher compared to the one obtained under Model b. Hence, model b could be considered for making inference.
The convergence of this model in Monolix is confirmed by looking at the SAEM convergence plot (the result is not shown). The appropriateness of the model is viewed by looking at the plot between individual observations v.s model prediction in Figure 8 and the individual fits plot shown for first 12 patients in Figure 13 (Appendix). It can be seen that the model fits data well.

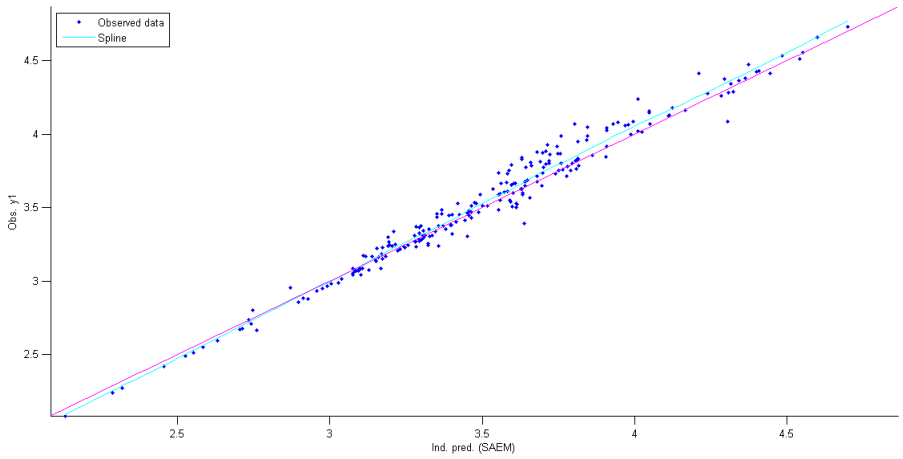


Figure 8: Individual observations v.s model prediction, both are presented on log_{10} scale (Reduced asymptotic model with covariates)

Parameter Estimation

Table 6 shows the parameter estimates with their standard errors obtained under the final reduced asymptotic model. Compared results from Table 6 and Table 5, it is seen that the estimates of b and A_0 are quite stable while the estimates of ϕ_s, μ_s and μ_A in reduced asymptotic model are lower compared to those obtained under asymptotic model. In the reduced asymptotic model, the lifespan of short-lived plasma cells is approximately five times longer to those of antibodies. This number is even smaller than the number

obtained under asymptotic model in which the lifespan of antibodies is nearly 9 times shorter than short-lived plasma cells. Hence, the assumption of $\mu_s << \mu_A$ in this case might not be reasonable.
VL significantly affects estimates of μ_A with a negative coefficient of -0.4990 ($p-value = 0.0310$) meaning those patients with higher VL might have smaller estimated antibody decay, i.e., having longer antibody lifespan. Both AGE and AV significantly affect the estimate of μ_s. Those patients were treated with antiviral drug would have lower estimated decay rate of short-lived plasma cells, leading to a longer lifespan of these cells ($\beta_{\mu_s,AV1} = -0.6410$). Reversely, patients are older would have on average larger estimates of μ_s meaning the lifespans of these cells are shorter ($\beta_{\mu_s,AGE} = 0.0227$). It is noticed that the older patients also have higher estimated ϕ_s, i.e, having either higher initial long-lived plasma population size (P_l^0) or higher average production rate of antibody from these cells.

Table 6: Parameter estimates and their standard errors (Reduced asymptotic model with three covariates AGE, VL, AV)

	Parm	**s.e. (lin)**	**r.s.e. (%)**	**p-value**
ϕ_s	562	260	46	-
μ_A	0.0281	0.0220	79	-
μ_s	0.0057	0.0018	31	-
b	2010	240	12	-
A_0	2040	360	18	-
ω_{ϕ_s}	0.0357	1.3000	3720	-
ω_{μ_A}	1.3900	0.1700	13	-
ω_{μ_s}	0.1400	0.1800	130	-
ω_b	0.6640	0.0830	13	-
ω_{A_0}	1.3500	0.1300	9	-
a	0.1070	0.0094	9	-
$\beta_{\mu_A,AGE}$	0.0414	0.0170	40	0.130
$\beta_{\mu_A,AV1}$	1.9900	1.3000	68	0.1400
$\beta_{\mu_A,tVL}$	-0.4490	0.2100	46	**0.0310**
$\beta_{\mu_s,AGE}$	0.0227	0.0066	29	**0.0006**
$\beta_{\mu_s,AV1}$	-0.6410	0.2600	40	**0.0120**
$\beta_{\mu_s,tVL}$	-0.0323	0.0740	231	0.6600
$\beta_{\phi_s,AGE}$	0.0313	0.0120	38	**0.0077**
$\beta_{\phi_s,AV1}$	1.1200	1.2000	107	0.3500
$\beta_{\phi_s,tVL}$	0.0395	0.1200	308	0.7500

5.3 Model to investigate the relationship between PHN and antibody titers

5.3.1 Random-Effects Model

Model with LOGAB as covariate

The proportional odds model is fitted using PROC NLMIXED in SAS. First of all, the model containing only $time, logab, time*logab$ as covariates is fitted with the assumption of proportional odds. At the next step, the $time^2$ effect is added to the model. The likelihood ratio test gives a very small $p - value$ (< 0.0001) indicating that the model with $time^2$ effect is preferable over the model without $time^2$ effect. Again, based on the likelihood ratio test, it is shown that the addition of the interaction term between $time^2$ and $logab$ to this model is not necessary. Moreover, the addition of $time^3$ effect is not important. The model 15 as specified in Section 4.3.2 is considered as the final one.
Given the mean structure specified as in Model 15, a non-proportional odds model and 14 partial proportional odds model (4 models which have parameters different in subset of 1 covariate, 6 models assuming parameters different in subset of any 2 covariates and 4 models having parameters different in subset of any 3 covariates) are fitted. Table 9 in the Appendix shows the $-2LogL$, AIC, and $p - value$ for the likelihood ratio test to see if the proportional odds model is preferable over non-proportional odds ratio and partial proportional odds model. It can be seen that models $(2), (3), (5), (8), (11)$ having $p - value < 0.05$ which indicates that these models are preferable over proportional odds model. Among the four models, model (8) has the smallest $p - value$. It is decided to use this model for making inference. The model could be written as:

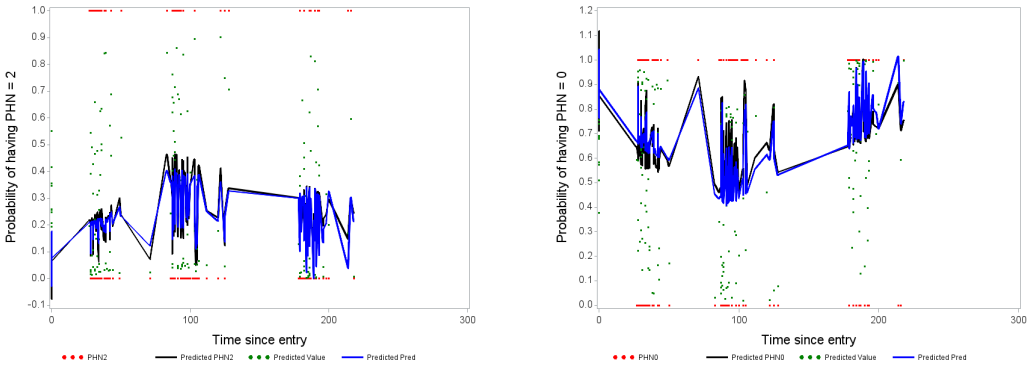
$$logit[P(PHN_{ij} \leq 0)] = \beta_{01} + b_i + \beta_{11}t_{ij} + \beta_2 LOGAB_{ij} + \beta_3 t_{ij} LOGAB_{ij} + \beta_{41}t_{ij}^2$$
$$logit[P(PHN_{ij} \leq 1)] = \beta_{02} + b_i + \beta_{12}t_{ij} + \beta_2 LOGAB_{ij} + \beta_3 t_{ij} LOGAB_{ij} + \beta_{42}t_{ij}^2$$

where b_i is the random intercept, accounting for the inter-individual variability. b_i is assumed to have normal distribution $b_i \sim N(0, \sigma_b^2)$.

Model diagnostics

Fitting the model in SAS using PROC NLMIXED, it gives AIC of 339.5. There is not many documentation about a common diagnostic methods for a GLMM model. In the context of this thesis, we use the method to compare between predicted probabilities and observed probabilities as a simple mean for model diagnostics. As the observed probabilities take only two values (either 0 or 1), we employ the smoothing technique in PROC LOESS in SAS where the probability of having $PHN = k$ $(k = 0, 1, 2)$ is smoothed overy $time$ and $logab$. The same technique is applied to prediction probabilities Of having $PHN = 0, 1, 2$. To serve the comparison purpose, it is decided to use the same smoothing parameter for two smoothing processes (the average of two optimal smoothing parameters when running two separate loess procedures). The smoothed surfaces of observed probabilities and predicted probabilities are compared. Figure 9 shows the smoothed predicted probabilities (blue line) and smoothed observed probabilities (black

line) for probability of having PHN, i.e., $P(PHN = 2)$ (left panel) and probability of having no pain, i.e., $P(PHN = 0)$. It is seen that the two lines almost overlap each other. The same observation is observed when looking at the plot of probabilities over different values of $LOGAB$ in Figure 14 (Appendix). It is concluded that the fit of the model could be appropriate. We use this model for making inference.



(a) $P(PHN = 2)$ (b) $P(PHN = 0)$

Figure 9: The smoothed lines of predicted probabilities and observed probabilities: The left panel describes the change of $P(PHN = 2)$ over time, the right panel shows the change of $P(PHN = 0)$ over time.

Parameter Estimation

The parameter estimates of this model is given in Table 7. All the effects are significant.

Table 7: Parameter estimates for partial proportional odds model (8)

Parm	Logit	Effect	Estimate	SE	p-value
β_{01}	(1)	Intercept	8.9393	2.8822	0.0029
β_{02}	(2)	Intercept	9.2387	2.9548	0.0027
β_{11}	(1)	$time$	-0.1583	0.0403	0.0002
β_{12}	(2)	$time$	-0.1262	0.0399	0.0025
β_2	(1) & (2)	$LOGAB$	-1.6632	0.7421	0.0287
β_3	(1) & (2)	$time * LOGAB$	0.0234	0.0091	0.0127
β_{41}	(1)	$time^2$	0.0004	0.0001	$< .0001$
β_{42}	(2)	$time^2$	0.0002	0.0001	0.0233
σ_b	(1) & (2)		2.2481	0.5106	$< .0001$

The two logits could be now written down:

$$\begin{aligned} logit[P(PHN_{ij} \leq 0)] &= -0.1583t_{ij} - 1.6632LOGAB_{ij} + 0.0234t_{ij}LOGAB_{ij} \\ &\quad + 0.0004t_{ij}^2 + b_i + 8.9393 \\ logit[P(PHN_{ij} \leq 1)] &= -0.1262t_{ij} - 1.6632LOGAB_{ij}0.0234t_{ij}LOGAB_{ij} \\ &\quad + 0.0002t_{ij}^2 + b_i + 9.2387 \end{aligned}$$

Since the effects of $time$, $time^2$, $LOGAB$, $time * LOGAB$ are all significant, it is a little bit cumbersome to interpret model's parameters. Let us consider those subjects at baseline, i.e., time point $t_{ij} = 0$ which do not vary from the population ($b_i = 0$) having $LOGAB = 3.5392$ (the mean value of $LOGAB$). The expected probability of having $PHN = 0$ is 96% (with 95% CI of approximately [90%; 100%]); the probability of having $PHN \leq 1$ (having no pain or dyskinesia) is 97% (95% CI of [91%; 100%]) and the probability of having PHN is very small of 3% (95% CI of [0; 8%]). This result is reasonable as we do not expect to see PHN at the baseline. Next, we consider those subjects at later time points, i.e., $t = 60, 120, 200$ respectively (equivalently as visit 2, 3 and 4), having $LOGAB = 3.5392$. The expected probabilities with their CIs are provided in Table 8. We notice that the probability of having PHN, i.e. $P(PHN = 2)$

Table 8: Expected probabilities and their 95% CI for those subjects with no deviation from the population, i.e., $b_i = 0$ at 4 different time points corresponding approximately to visit 1, 2, 3 and 4 provided that $LOGAB = 3.5392$

Time	Probability	Estimate	SE	Lower	Upper
t = 0	$P(PHN = 0)$	0.9549	0.0293	0.8964	1.0134
	$P(PHN \leq 1)$	0.9662	0.0273	0.9117	1.0207
	$P(PHN = 2)$	0.0338	0.0273	-0.0207	0.0883
t = 60	$P(PHN = 0)$	0.4755	0.1072	0.2611	0.6899
	$P(PHN \leq 1)$	0.8189	0.0717	0.6755	0.9623
	$P(PHN = 2)$	0.1811	0.0717	0.0377	0.3245
t = 120	$P(PHN = 0)$	0.3742	0.1093	0.1555	0.5928
	$P(PHN \leq 1)$	0.7627	0.1041	0.5544	0.9709
	$P(PHN = 2)$	0.2373	0.1041	0.0291	0.4456
t = 200	$P(PHN = 0)$	0.9602	0.02905	0.9021	1.0183
	$P(PHN \leq 1)$	0.9547	0.03525	0.8842	1.0252
	$P(PHN = 2)$	0.0453	0.035	-0.0252	0.1158

is higher at approximately days 60 and 120 (corresponding to visit 2 and 3). However, it is worth emphasizing that these probabilities also depend on the magnitude of $LOGAB$.

Model with additional covariates: AGE, AV

The model with covariates is fitted in SAS using PROC NLMIXED. Firstly, only covariates AGE, AV are considered without the interaction with time. The partial proportional odds model (taking the same form as in the model with only $LOGAB$ as covariate) shows non significant effects of both age and antiviral usage. The inclusion of interaction term between these two covariates and time into the model also gives non significant $p-values$. It is concluded that the effects of age and antiviral usage on the probability of having PHN are not significant.

6 Conclusion and Discussion

In this thesis, the imprinted lifespan model is employed to study the longitudinal dynamics of HZ. The model represents the actual process including three time-scales (antibody, long-lived plasma cells and short-lived plasma cells) (Amanna and Slifka, 2010). It turns out that the asymptotic model proposed by Andraud et. al. (2012) shows a good fit to the data. However, the model could not estimate the production rates of antibodies by short- and long-lived plasma cells, i.e., φ_s, φ_l due to its formulation. It is only able to estimate the quantity ϕ_s, ϕ_l which could be interpreted as the number of antibodies produced by short- and long-lived plasma cells at beginning. In the limitation of this thesis, only estimated standard errors obtained with model linearization are presented. In addition to these numbers, bootstrap methods can also be proceeded. This technique is useful for estimating the distribution of statistics without using asymptotic theory and with very few assumptions on the data distribution (Lavielle, 2015). This has not been done but it could be applied for further analysis.
The inclusion of covariates AGE, VL, AV improves the fit of the model. Viral load at the beginning of the study and antiviral treatment significantly influence the individual estimates of μ_A. Those patients on antiviral treatment have higher estimated antibody decay rate, i.e., shorter antibody lifespan (while keeping other factors as fixed) compared to those were not prescribed antiviral drug. Age of patients at entry and antiviral treatment are important in estimating the individual estimates of μ_s. If the viral loads at baseline are higher, the estimated antibody and short-lived plasma cells decay rates are lower. This leads to longer antibody and short-lived plasma cells lifespan. Older patients have lower estimated short-lived plasma cells decay meaning shorter lifespan of these cells. Viral load data are collected repeatedly at each measurements, hence a non-linear mixed model with two ouputs (antibody level and viral load) or a model considering viral load as design variable could be considered. These approaches will take all the information of viral load into account.
As stated earlier in the Methodology part, the approach of Andraud et al. (2012) might be not appropriate to our data since we have continuous exposure. However, we see that only by assuming the decay rate of long-lived plasma cells could be ignored, the model fits our data well. One possible explanation could be that some patients came to the clinic quite late after having symptoms of HZ. For those patients, their antibodies are highest at the first measurement and decline over time.
The reduced asymptotic model also shows a good fit. However, given the fact that this model put more constraints on the parameters, it might have the limitation to generate the method to analysis of other diseases or processes.
GLMM approach is used to investigate the relationship between having PHN at each measurement and antibody levels. The random intercept partial proportional odds model shows that $time, time^2, LOGAB$ significantly affect the probability of having PHN. Generally, one might expect that the probability of having PHN ($P(PHN = 2)$) is low at the very beginning of the disease, higher at the middle period and decreasing again in very late time. The inclusion of two covariates AGE, AV shows that they are not significantly influencing the probability of having PHN. This result shows inconsistent

conclusion against claims that age and antiviral usage play important roles in the appearance and control of PHN in HZ patients. There is no solid explanation for this. One possible reasoning is that age and antiviral drug treatment possibly influence antibody levels in human body. Since the model takes antibody titers at each measurements into account, consequently, age and antiviral treatment no longer play vital role in predicting the probability of having PHN.

A Generalized Estimating Equations (GEE) model could be employed to investigate the population relationship between antibody titers and having PHN . This model takes into account the inter-individual variability in a repeated measurements study and could be fitted in SAS using GENMOD procedure. However, If the response of interest is a multi-categorical variable, there is only one choice of independent working correlation for proportional odds model (Molenberghs and Verbeke, 2005). Furthermore, if the proportional odds assumption is not satisfied, it is not possible to fit non-proportional odds or partial proportional odds model by using marginal approach. Since the partial proportional model is shown to be the most appropriate one in this study, the marginal approach is not considered.

The data set contains missing values for AB. Our non-linear mixed effects models are fitted within Monolix which uses SAEM algorithm to estimate the maximum likelihood for population parameters. SAEM algorithm is a stochastic approximation of EM algorithm which was developed initially to estimate models with missing or non-observed data such as random effects (Panhard and Samson, 2009). Hence, using this algorithm, the missing data could be deal with, the analysis still gives valid estimates.

For GLMM model, we report results from fitting the model with original data (the data with missing observations). Later on, we use the model developed in the first part (model relates AB on log_{10} scale with $time$) to predict antibody titers for those missing data points. GLMM mode is refitted with imputed dataset. The results show that parameter estimates, standard errors and $p-values$ between the two fits are very close. The GLMM model used could be viewed as less sensitive to those missing values. The inference made from this model could be valid.

Appendix

Figures

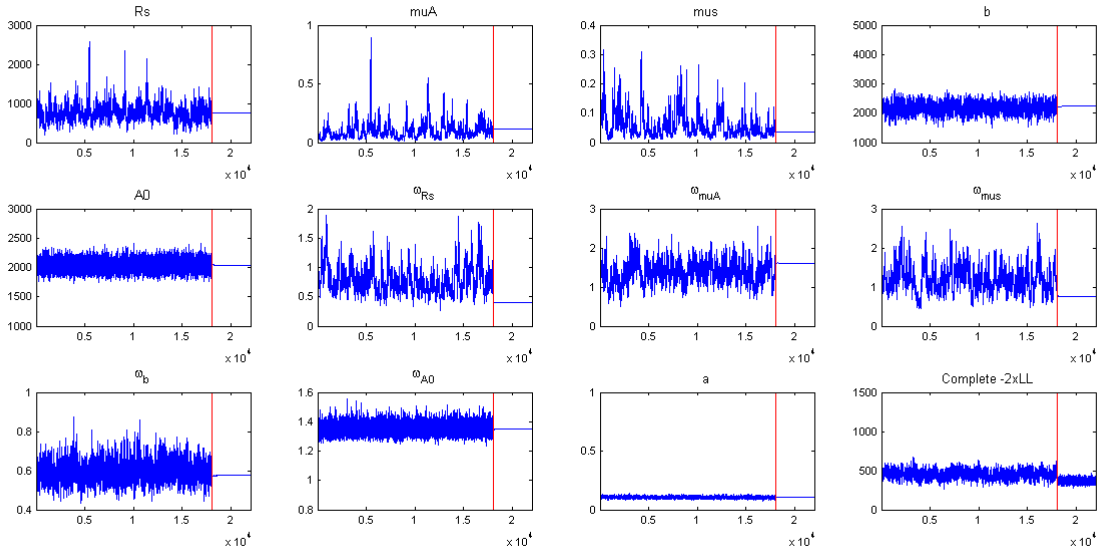


Figure 10: Convergence of SAEM (Asymptotic model without covariates). The vertical dotted line indicates where the algorithm switches from the first phase to the second.

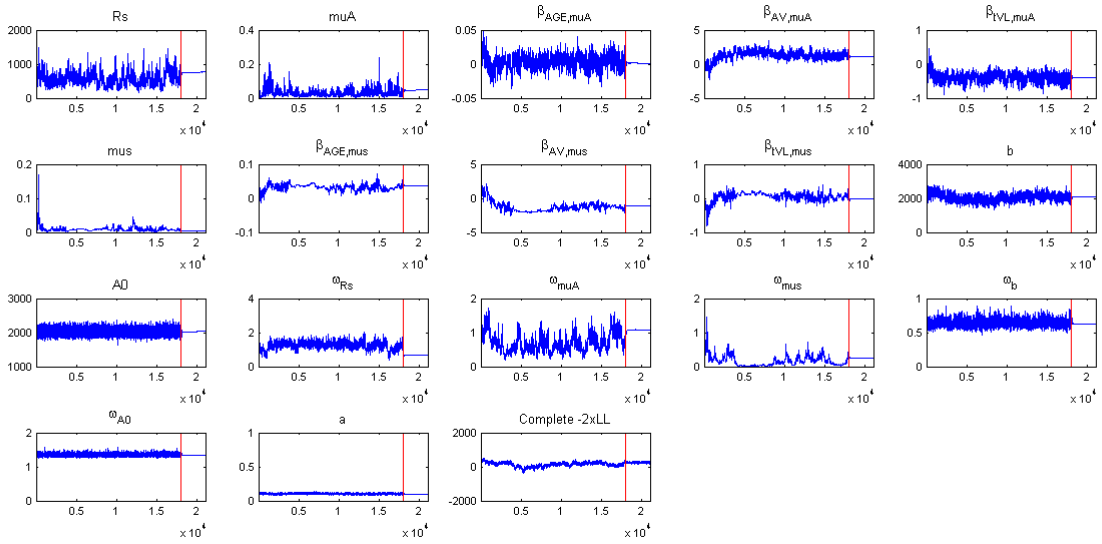


Figure 11: Convergence of SAEM (Asymptotic model with three covariates: AGE, VL, AV). The vertical dotted line indicates where the algorithm switches from the first phase to the second.

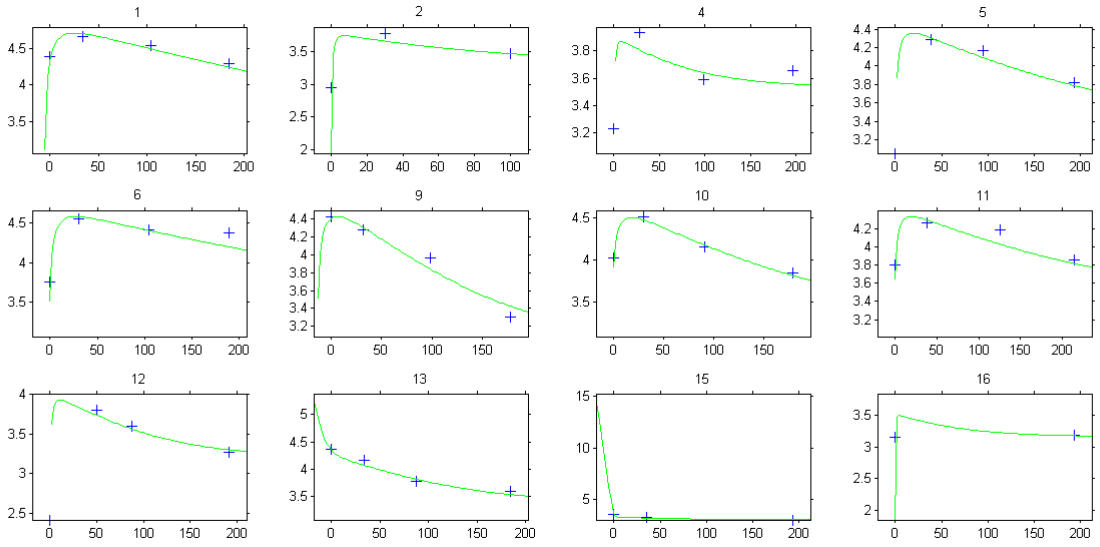


Figure 12: Individual fits for first 12 patients (Reduced asymptotic model without covariates).

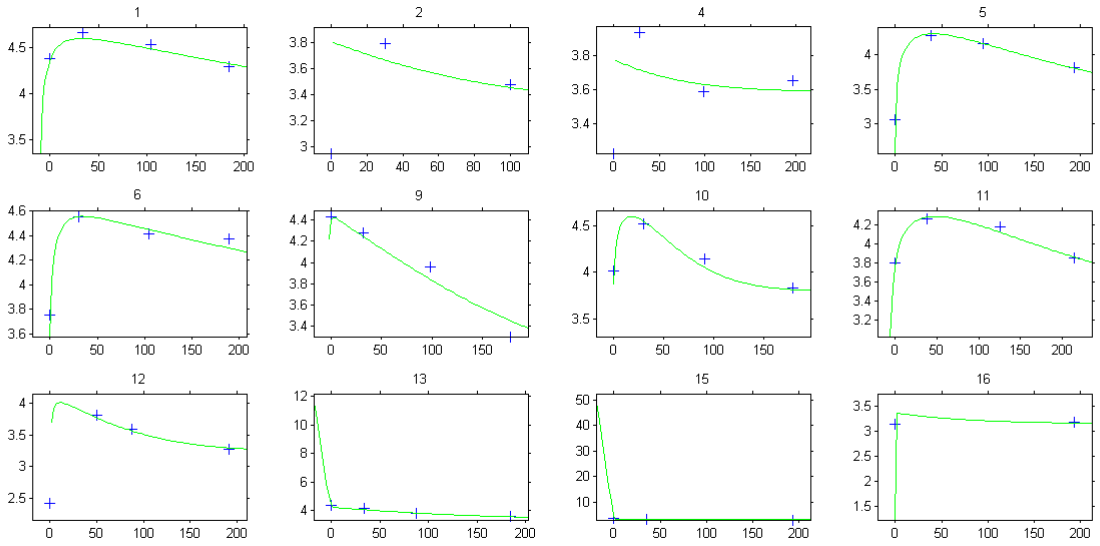
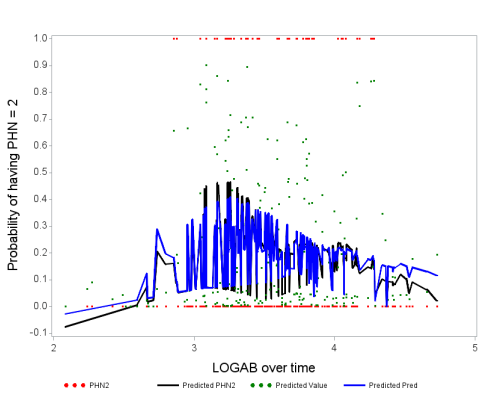
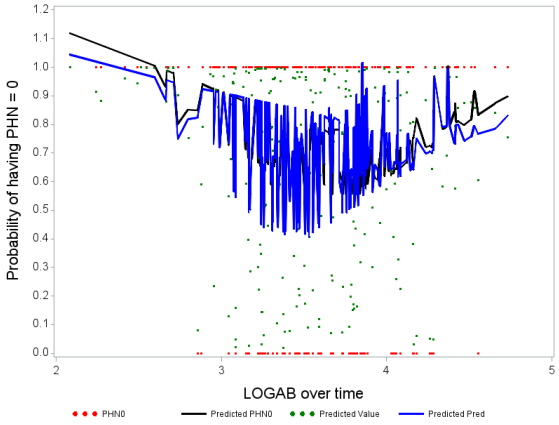


Figure 13: Individual fits for first 12 patients (Reduced asymptotic model with covariates)



(a) $P(PHN = 2)$



(b) $P(PHN = 0)$

Figure 14: The smoothed lines of predicted probabilities and observed probabilities: The left panel describes the change of $P(PHN = 2)$ with different values of $LOGAB$, the right panel shows the change of $P(PHN = 0)$ with different values of $LOGAB$.

Tables

Table 9: Models information for non-proportional odds and partial proportional odds models

	Model	**-2LogL**	**-2deltaL**	**DF**	**p-value**	**AIC**
Non-PO	(0)	320.6	8.8	4	0.0663	342.6
Same parameter in subset of 1 covariate						
β_1	(1)	323.1	6.3	3	0.0979	343.1
β_2	(2)	321.3	8.1	3	**0.0439**	341.3
β_3	(3)	321.4	8	3	**0.0460**	341.4
β_4	(4)	325.3	4.1	3	0.2509	345.3
Same parameters in subset of 2 covariates						
β_1, β_2	(5)	323.1	6.3	2	**0.0429**	341.1
β_1, β_3	(6)	324.4	5	2	0.0821	342.4
β_1, β_4	(7)	325.6	3.8	2	0.1496	343.6
β_2, β_3	(8)	321.5	7.9	2	**0.0195**	339.5
β_2, β_4	(9)	325.7	3.7	2	0.1572	343.7
β_3, β_4	(10)	325.7	3.7	2	0.1572	343.7
Same parameters in subset of 3 covariates						
$\beta_1, \beta_2, \beta_3$	(11)	324.5	4.9	1	**0.0269**	340.5
$\beta_1, \beta_2, \beta_4$	(12)	325.7	3.7	1	0.0544	341.7
$\beta_2, \beta_3, \beta_4$	(13)	325.7	3.7	1	0.0544	341.7
$\beta 1$,β_3, β_4	(14)	328.1	1.3	1	0.2542	344.1

References

Agresti, A. (2002). *Categorical Data Analysis, Second Edition*, Wiley.

Amanna, I. and Slifka, M. (2010). Mechanisms that determine plasma cell lifespan and the duration of humoral immunity, *Immunol Rev.* **236**: 125–138.

Andraud, M., Lejeune, O., Musoro, J., Ogunjimi, B., Beutels, P. and Hens, N. (2012). Living on three time scales: The dynamics of plasma cell and antibody populations illustrated for hepatitis a virus, *PLoS Comput Biol.* **8**(3).

Bader, M. (2013). Herpes zoster: diagnostic, therapeutic, and preventive approaches., *Postgrad Med.* **125**(5): 78–91.

CDC (2015). Shingles (herpes zoster).
URL: *http://www.cdc.gov/shingles/about/overview.html.*

Dworkin, R. and Schmader, K. (2001). Epidemiology and natural history of herpes zoster and post-herpetic neuralgia. in: Watson cpn, editor. herpes zoster and postherpetic neuralgia. amsterdam (netherlands), *New York: El-sevier Press.* pp. 39–64.

Dworkin, R. and Schmader, K. (2003). Treatment and prevention of post-herpetic neuralgia., *Clinical Infectious Diseases.* **36**: 877–82.

Gerhson, A. (1996). Epidemiology and management of postherpetic neuralgia., *Semin Dermatol.* **15 (2 Suppl)**(8-13): 274–280.

Hedeker, D. and Gibbons, R. (2006). *Longitudinal Data Analysis*, Wiley.

Karlsson, M. and Savic, R. (2007). Diagnosing model diagnostics, *Clin Pharmacol Ther.* **82**: 17–20.

Lavielle, M. (2015). *Mixed effects models for the Population Approach: Models, Tasks, Methods and Tools*, CRC Press, United States of America.

Lee, T., Hayes, S., Cummings, D., Cao, Q., Carpenter, K., Heim, L. and Edwards, H. (2013). Herpes zoster knowledge, prevalence, and vaccination rate by race, *J Am Board Fam Med.* **26**(1): 45–51.

Lixoft (2014). Monolix methodology: Version 4.3.2, a software for the analysis of nonlinear mixed effects models.
URL: *http://download.lixoft.com/data/packages/mlx-4.3.2/guides/monolixMethodology.pdf*

Martini, F. and Nath, J. (2011). *Fundamentals of Anatomy and Physiology*, Pearson Education, Inc., United States of America.

Molenberghs, G. and Verbeke, G. (2005). *Models for discrete longitudinal data*, Springer, New York.

Panhard, X. and Samson, A. (2009). Extension of the saem algorithm for nonlinear mixed models with 2 levels of random effects, *Biostatistics.* **10**(1): 121–135.

Pinchinat, S., Cebrian-Cuenca, A., Bricout, H. and Johnson, R. (2013). Similar herpes zoster incidence across europe: results from a systematic literature review, *BMC Infect Dis.* **13**(170).

Sampathkumar, P., Drage, L. and Martin, D. (2009a). Herpes zoster (shingles) and post-herpectic neuralgia., *Mayo Clin Proc.* **84(3)**: 274–280.

Sampathkumar, P., Drage, L. and Martin, D. (2009b). Herpes zoster (shingles) and postherpetic neuralgia, *Mayo Clin Proc.* **84**(3): 274–280.

Siegrist, C. (2015). Vaccine immunology.
URL: *http://www.who.int/immunization/documents/Elsevier_Vaccine_immunology.pdf*

Whitley, R., Weiss, H., Soong, S. and et, a. (1999). Herpes zoster: risk categories for persistent pain., *J Infect Dis.* **179**: 9–15.

Codes

R Code for Data Management: Data without PHN

```
#***************************** MANAGEMENT WORK ************************#
setwd("D:\\MASTER PROGRAMME\\...\\MY THESIS\\Data And Analysis\\Data")
getwd()

#***************************** READ IN DATA ************************#
data  <- read.table(file.choose(), sep = "\t", na.strings = ".",
                      header = TRUE) # Choose file: data_MPTT
str(data)
dim(data)
range(data$AGE)
table(data$AGE)
data$AGEcls <- NULL
data$AGEcls[data$AGE <= 50] <- 0
data$AGEcls[data$AGE > 50]  <- 1
table(data$AGEcls)

#View(data)
#write.csv(data,file="dataSAS.csv", row.names = F)
   # Check missing data according to AB
miss.AB <- which(is.na(data$AB) == TRUE )
miss.AB
miss.data.AB <- data[which(is.na(data$AB)==TRUE),]
miss.data.AB
   # Check missing data according to VL
miss.VL <- which(is.na(data$VL == TRUE))
miss.VL
miss.data.VL <- data[which(is.na(data$VL) == TRUE),]
miss.data.VL
   # How many patients
check <- unique(data$ID)
check
length(check)
   # Check frequency
freq <- data.frame(table(data$ID))
View(freq)
freq1 <- freq[which(freq[,2]==1),]
freq1
freq2 <- freq[which(freq[,2]==2),]
freq2
freq3 <- freq[which(freq[,2]==3),]
```

```
freq3

# Need some
vl <- data$VL[complete.cases(data$VL)]
sdvl <- sd(vl); sdvl
meanvl <- mean(vl); meanvl
age <- data$AGE[complete.cases(data$AGE)]
meanage <- mean(age); meanage
sdage <- sd(age); sdage
```

R Code for Data Management: Data with data on PHN

```
# Read in data: Read file PHN.csv
setwd("D:\\MASTER PROGRAMME\\...\\MY THESIS\\Data And Analysis\\PHN_Analysis")
phn2 <- read.csv(file.choose(),header = T, sep = ",",dec = ".",na.strings="#NULL!")
str(phn2)
library(plyr)
phn2= rename(phn2,c("IGG_V1"="ABV1","IGG_V2"="ABV2","IGG_V3"="ABV3","IGG_V4"="ABV4",
   "LOGIGGV1"="LOGV1","LOGIGGV2"="LOGV2","LOGIGGV3"="LOGV3","LOGIGGV4"="LOGV4",
   "V2STATUS"="PHNV2","V3STATUS"="PHNV3","V4STATUS"="PHNV4"))
str(phn2)

# Reshape the data
phn3 <- phn2[,-c(6,7,8,9)]
str(phn3)
    #View(phn3)
phn <- reshape(phn3,direction="long",varying=c("ABV1","ABV2","ABV3","ABV4"),
v.names = "AB", idvar = "ID",timevar = "visit", times = c(1,2,3,4))
str(phn)
    #View(phn)
phn <- phn[order(phn$ID),]
    #View(phn)
# Save the file in CSV format
write.csv(phn,row.names = F, file = "phn2.csv")

# Create the data with corresponding PHNV status
phnV2_0 <- phn[phn$PHNV2 ==0,]
str(phnV2_0) # 268 observations

phnV3_0 <- phn[phn$PHNV3 ==0,]
str(phnV3_0)  # 264 observation
phnV3_1 <- phn[phn$PHNV3 == 1,]
```

```
str(phnV3_1) # 0 observations
phnV3_3 <- phn[phn$PHNV3 == 3,]
str(phnV3_3) # 4 observations

phnV4_0 <- phn[phn$PHNV4 == 0,]
str(phnV4_0) # 0 observation
phnV4_1 <- phn[phn$PHNV4 == 1,]
str(phnV4_1) # 260 observation
phnV4_3 <- phn[phn$PHNV4 == 3,]
str(phnV4_3) # 8 observations

# Remove those subjects NOT having measurements of AB at all out of the data set
# so as to we have the same data set with part I
   # View(phn)
phn_same <- phn[which(phn$ID != 3 & phn$ID !=8 & phn$ID != 22 &
phn$ID != 33 & phn$ID != 40 &phn$ID != 42),]
   # View(phn_same)
str(phn_same)
str(phn)
phn_ab <- write.csv(phn_same, file = "phn_ab.csv", row.names = F, quote = F)
# Read in the data file from the first part
data <- read.csv(file.choose(),header = T, sep = ",",dec = ".",na.strings="#VALUE!")
str(data)
```

SAS Code for Data Exploratory

```
/*READ IN DATA*/
data initial;
   infile 'D:\MASTER PROGRAMME\...\Data\dataSAS.csv' firstobs = 2 dlm = ",";
   input ID age gen av dur time ab cmv vl;
run;
proc print data = initial;
run;

/*EXPLORE DATA*/
/*Simple statistics summary*/
proc contents data = initial;
run;
proc means data = initial;
   var ab vl;
run;
```

```
/*Explore the individual profile*/
goptions reset = all ftext = swiss device = psepsf gsfname =fig0 gsfmode = replace
rotate = landscape i = join;
proc gplot data = initial;
    plot ab*time = ID / haxis = axis1 vaxis = axis2;
axis1 label=(h=2 "Time since entry") value=(h=1.5) minor=none order=(0 to 250 by 50);
axis2 label = (h=2 "AB level") value = (h=1.5) minor =none;
run;
    /*Individual profile of a subset*/
data id; set initial; by ID; if last.ID then output; proc print; run;
data toselect; set id; proc print; run;

proc sql OUTOBS=30 ;
    create table subID as
    select A.*
    from toselect as A
    order
    by RANUNI(4537) ;
quit;

proc print data = subID; run;
data sub; set initial; where ID in (4 50 13 20 69 73 65 31 48 30 59 38 1
                    53 24 15 63 11 46 2 6 21 68 49 45 71 56 61 70 43);
proc print; run;
goptions reset = all i = join;
proc gplot data = sub;
    plot ab*time = ID / haxis = axis1 vaxis = axis2;
axis1 label=(h=2 "Time since entry") value=(h=1.5) minor=none order=(0 to 250 by 50);
axis2 label = (h=2 "AB level") value = (h=1.5) minor = none;
run;

/*Missing: 2 missings of vl, 4 missings of ab */
/*Unbalanced data: Use smoothed loess*/
/*Explore The Mean Structure - Unbalance Data*/
/*Calculate mean for each time point with their s.e*/
proc loess data = initial;
    ods output scoreresults = out;
model ab = time;
score data = initial;
run;
proc sort data = out; by time; run;
proc print data = out; run;
proc means data = out;
    var p_ab;
run;
```

```
goptions reset = all ftext = swiss rotate = landscape;
proc gplot data = out;
    plot ab*time = 1 p_ab*time = 2 / overlay haxis = axis1 vaxis = axis2;
    symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
    axis1 label = (h=2 "Time since entry") value=(h=1.5)
    minor=none order=(0 to 250 by 50);
axis2 label = (h=2 A=90 "AB values") value=(h=1.5)  minor = none;
run;
/*Transform AB into log10(AB)*/
data initial; set initial; logab = log10(ab); run; proc print; run;
proc loess data = initial;
    ods output scoreresults = out2;
model logab = time;
score data = initial;
run;
proc sort data = out2; by time; run;
proc print data = out2; run;
proc means data = out2;
    var p_logab;
run;
proc gplot data = out2; /*Transform AB to log(AB)*/
    plot logab*time = 1 p_logab*time = 2 / overlay haxis = axis1 vaxis = axis2;
    symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
    axis1 label = (h=2 "Time since entry") value=(h=1.5)
    minor=none order = (0 to 250 by 50);
axis2 label = (h=2 A=90 "log10(AB) values") value=(h=1.5)  minor = none;
run;
```

Monolix Code

```
    DESCRIPTION:
Complete Model - 2 ODEs

INPUT:
parameter = {Rs,muA,mus,b,A0}

EQUATION:
A = (Rs/(muA-mus))*exp(-mus*t) + b + (A0 - Rs/(muA-mus)-b)*exp(-muA*t)
AB = log10(A)
```

```
OUTPUT:
output = {AB}
```

SAS Code: GLMM analysis

```
data phn;
   infile "D:\MASTER PROGRAMME\...\phn_part2_ok.csv" dlm = "," firstobs = 2;
   input AB ID TIME PHN VL PHNV2 PHNV3 PHNV4 VISIT LOGAB;
run;                                          /*Fit PO model for PHN*/
/*TIMECLASS*/
data phn_time; set phn; timecls = time; TIMELOG = time*LOGAB;
time2 = time**2;time3 = time**3; run;
/*Combine output 0 and 1*/
data phn_com; set phn_time; if PHNV2 = 0 | PHNV2 = 1 then V2STATUS = 0;
                      else if  PHNV2 = 2 then V2STATUS = 1; else V2STATUS = PHNV2;
   if PHNV3 = 0 | PHNV3 = 1 then V3STATUS = 0;
   else if PHNV3 = 2 then V3STATUS = 1; else V3STATUS = PHNV3;
   if PHNV4 = 0 | PHNV4 = 1 then V4STATUS = 0;
   else if PHNV4 = 2 then V4STATUS = 1; else V4STATUS = PHNV4;
run;
/*GEE*/
proc genmod data = phn_com;
    class ID timecls;
model PHN = time LOGAB time*LOGAB / dist = multinomial link = cumlogit;
repeated subject = ID/ type = ind covb corrw within = timecls modelse;
run;  /*QIC = 397.9084. All are significant*/
proc genmod data = phn_com;
    class ID timecls;
model PHN = time LOGAB time*LOGAB time2 / dist = multinomial link = cumlogit;
/*Model with time^2 term*/
repeated subject = ID/ type = ind covb corrw within = timecls modelse;
run;  /*QIC = 374.1342. LOGAB is not significant*/
proc genmod data = phn_com;
    class ID timecls;
model PHN = time LOGAB time*LOGAB time2 time2*LOGAB /
        dist = multinomial link = cumlogit;
/*Model with time^2 & time^2 * LOGAB term*/
repeated subject = ID/ type = ind covb corrw within = timecls modelse;
run; /*Don't use this model*/
proc genmod data = phn_com;
    class ID timecls;
model PHN = time LOGAB time*LOGAB time2 time3 / dist = multinomial link = cumlogit;
```

```
/*Model with time^2 & time^3*/
repeated subject = ID/ type = ind covb corrw within = timecls modelse;
run;  /*Don't use this model*/

/*GLMM*/
/*Mixed model: Random intercept + Random slope*/
proc glimmix data = phn_com method = RSPL;
    title "PROC GLIMMIX analysis, ordinal, RSPL (PQL, REML)";
class timecls id; /*PO: Not put cat2 into class statement*/
nloptions maxit = 100;
model phn = time LOGAB time*LOGAB / dist = multinomial link = cumlogit solution;
random intercept time / subject = id type = un;
/*Random intercept and slope*/
/*Variance structure: unstructure: DID NOT CONVERGE*/
run;


proc glimmix data = phn_com method = RSPL;
    title "PROC GLIMMIX analysis, ordinal, RSPL (PQL, REML)";
class timecls id; /*PO: Not put cat2 into class statement*/
nloptions maxit = 50;
model phn = time LOGAB time*LOGAB/ dist = multinomial link = cumlogit solution;
random intercept time / subject = id type = cs;
/*Random intercept and slope*/
/*Variance structure: DID NOT CONVERGE*/
run;


/*Model: Random Intercept only*/
proc glimmix data = phn_com method = RSPL ;
    title "PROC GLIMMIX analysis, ordinal, RSPL (PQL, REML)";
class timecls id; /*PO: Not put cat2 into class statement*/
nloptions maxit = 50;
model phn = time LOGAB time*LOGAB / dist = multinomial link = cumlogit solution ;
random intercept / subject = id type = un;
/*Random intercept */
/*Variance structure: unstructure*/
run;  /*All are significant, except for LOG*AB, -2L1 = 1543.60*/
proc glimmix data = phn_com method = RSPL ;
    title "PROC GLIMMIX analysis, ordinal, RSPL (PQL, REML)";
class timecls id; /*PO: Not put cat2 into class statement*/
nloptions maxit = 50;
model phn = time LOGAB time*LOGAB time2/
         dist = multinomial link = cumlogit solution ;
random intercept / subject = id type = un;
```

```
/*Random intercept */
/*Variance structure: unstructure*/
run;  /*All are significant, -2L2 = 1712.45*/
/*-2(L2-L1) = -2L2 + 2L1 = 1712.45 - 1543.60 = 168.85*/
    /*Model with interaction between time2 and logab*/
proc glimmix data = phn_com method = RSPL;
    class timecls id;
nloptions maxit = 50;
model phn = time logab time*logab time2 time2*logab/
        dist = multinomial link = cumlogit solution;
random intercept / subject = id type = un;
run; /*-2L = 1757.44. This model does not improve fit*/

    /*Model with time3 term*/
proc glimmix data = phn_com method = RSPL;
    class timecls id;
nloptions maxit = 50;
model phn = time logab time*logab time2 time3/
        dist = multinomial link = cumlogit solution;
random intercept / subject = id type = un;
run; /*-2L = 1994.96. This model does not improve fit*/

             /**************** USING NLMIXED ******************/
/*Assume the same slope*/
proc nlmixed data = phn_com qpoints = 20 TECH = NEWRAP;
  title 'Herpes Zoster Data, Proc Nlmixed, ordinal, adaptive, q = 20';
  parms beta1 = -0.0433 beta2 = -0.8194 beta3 = 0.0118 int1 = 3.7085
        int2 = 4.2871 d = 2;
  eta = beta1*time + beta2*LOGAB + (b2 + beta3)*time*LOGAB + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta - int2) + 1) - 1/(exp(-eta - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 b2 ~ normal([0,0],[d1*d1, d12, d2*d2]) subject = id; /*Unstructured*/
  estimate "var1" d1*d1;
  estimate "cov" d12;
  estimate "var2" d2*d2;
run;  /*Optimization can not be completed*/

/*Fixed effects model to get initial values*/
proc glimmix data = phn_com method = RSPL;
    title "PROC GLIMMIX analysis, ordinal, RSPL (PQL, REML)";
```

```
class timecls id; /*PO: Not put cat2 into class statement*/
nloptions maxit = 50;
model phn = time LOGAB time*LOGAB time2/
        dist = multinomial link = cumlogit solution;
run;

/* Model assume PO */
proc nlmixed data = phn_com qpoints = 20; /*If specify NOAD, almost same result*/
  title 'VZV Data, Proc Nlmixed, ordinal, adaptive, q = 20';
  parms beta1 = -0.0433 beta2 = -0.8194 beta3 = 0.01177 int1 = 3.7085
        int2 = 4.2871 d = 2;
  eta = beta1*time + beta2*LOGAB + beta3*TIMELOG + b1;
  /*random intercept*/
  if phn = 0 then z = 1/(exp(-eta - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta - int2) + 1) - 1/(exp(-eta - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var1" d*d;
run; /* -2logL = 364.3*/

/*Model with time^2*/
proc nlmixed data = phn_com qpoints = 20; /*If specify NOAD, almost same result*/
  title 'VZV Data, Proc Nlmixed, ordinal, adaptive, q = 20';
  parms beta1 = -0.09716 beta2 = -0.6936 beta3 = 0.01573 beta4 = 0.000207
       int1 = 4.1650
        int2 = 4.8017 d = 2;
  eta = beta1*time + beta2*LOGAB + beta3*TIMELOG + beta4*time2 + b1;
  /*random intercept*/
  if phn = 0 then z = 1/(exp(-eta - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta - int2) + 1) - 1/(exp(-eta - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var1" d*d;
run; /* -2logL = 329.4*/
/*All are significant*/

/*Assume different slopes - Non PO model*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
```

```
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta22 = -0.6936 beta32 = 0.01573 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta22*LOGAB + beta32*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run;  /* -2logL = 320.6, -2DeltaL = 8.8. DF = 4*/
/*LR Test: p-value = 0.06629764 > 0.05*/

/*Assume different slopes - Partial PO model (same parm in subset of 1 covariate).
DF = 3*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta22 = -0.6936 beta32 = 0.01573 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta11*time + beta22*LOGAB + beta32*time*LOGAB + beta42*time2 + b1;
  /*beta11 = beta12*/
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run;  /* -2logL = 323.1, -2DeltaL = 329.4- 323.1 = 6.3. DF = 3
p-value =  0.09789265*/

proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716  beta32 = 0.01573 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta21*LOGAB + beta32*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
```

```
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run;
 /*-2logL =321.3. -2DeltaL = 329.4-321.3 = 8.1. DF = 3
p-value = 0.04398959*/

proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta22 = -0.6936  beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta22*LOGAB + beta31*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 321.4. -2DeltaL = 329.4-321.4 = 8. DF = 3
p-value = 0.046*/

proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta22 = -0.6936 beta32 = 0.01573;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta22*LOGAB + beta32*time*LOGAB + beta41*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run;  /*-2LogL = 325.3. -2DeltaL = 329.4-325.3 = 4.1
```

```
p-value = 0.2509*/

/*Assume same parameters in subset of 2*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta32 = 0.01573 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta11*time + beta21*LOGAB + beta32*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run;  /*-2LogL = 323.1. -2DeltaL = 329.4-323.1 = 6.3. DF = 4-2 = 2
p-value = 0.0429*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta22 = -0.6936 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta11*time + beta22*LOGAB + beta31*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 324.4. -2DeltaL = 329.4-324.4 = 5. DF = 2
p-value = 0.0821*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta22 = -0.6936 beta32 = 0.01573 ;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta11*time + beta22*LOGAB + beta32*time*LOGAB + beta41*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
```

```
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 325.6. -2DeltaL = 329.4 - 325.6 = 3.8. DF =2
p-value = 0.1496*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta21*LOGAB + beta31*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 321.5. -2DeltaL = 329.4-321.5 = 7.9
p-value = 0.0195*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta32 = 0.01573;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta21*LOGAB + beta32*time*LOGAB + beta41*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 325.7. -2DeltaL = 329.4-325.7 = 3.7
p-value = 0.1572*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
```

```
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta22 = -0.6936;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta22*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 325.7. -2DeltaL = 3.7.
p-value = 0.1572*/

/*Parameters different in subset of 3. DF = 1*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 324.5. -2DeltaL = 329.4-324.5 = 4.9. DF = 1
p-value = 0.0269*/

proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta32 = 0.01573;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta11*time + beta21*LOGAB + beta32*time*LOGAB + beta41*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
```

```
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 325.7. -2DeltaL = 329.4-325.7 = 3.7
p-value = 0.0544*/

proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2LogL = 325.7. -2DeltaL = 329.4=325.7 = 3.7
p-value = 0.0544*/

proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta22 = -0.6936;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta11*time + beta22*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id;
  estimate "var" d*d;
run; /*-2Logl = 328.1. -2DeltaL = 329.4-328.1 = 1.3
p-value = 0.2542*/
```

```
                        /*MODEL FOR INFERENCE*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta21*LOGAB + beta31*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id out = random;
  estimate "var" d*d;
  /*At baseline*/
  estimate "P(0)" exp(int1+beta21*3.5392)/(1+exp(int1+beta21*3.5392));
  estimate "P(<= 1)" exp(int2+beta21*3.5392)/(1+exp(int2+beta21*3.5392));
  /*3.5392 = mean(logab)*/
  estimate "P(2)" 1 - exp(int2+beta21*3.5392)/(1+exp(int2+beta21*3.5392));
  /*At t = 60*/
  estimate "P(0)2" exp(int1+beta11*60+beta21*3.5392+beta31*60*3.5392+beta41*60*60)/
  (1+exp(int1+beta11*60+beta21*3.5392+beta31*60*3.5392+beta41*60*60));
  estimate "P(<=1)2" exp(int2+beta12*60+beta21*3.5392+beta31*60*3.5392+beta42*60*60)/
  (1+ exp(int2+beta12*60+beta21*3.5392+beta31*60*3.5392+beta42*60*60));
  estimate "P(2)2" 1-exp(int2+beta12*60+beta21*3.5392+beta31*60*3.5392+beta42*60*60)/
  (1+ exp(int2+beta12*60+beta21*3.5392+beta31*60*3.5392+beta42*60*60));
  /*At t = 120*/
  estimate "P(0)3"exp(int1+beta11*120+beta21*3.5392+beta31*120*3.5392+beta41*120*120)
  /(1+exp(int1+beta11*120+beta21*3.5392+beta31*120*3.5392+beta41*120*120));
  estimate "P(<=1)3" exp(int2+beta12*120+beta21*3.5392+beta31*120*3.5392+beta42*120*1
  /(1+exp(int2+beta12*120+beta21*3.5392+beta31*120*3.5392+beta42*120*120));
  estimate "P(2)3"1-exp(int2+beta12*120+beta21*3.5392+beta31*120*3.5392+beta42*120*12
  /(1+ exp(int2+beta12*120+beta21*3.5392+beta31*120*3.5392+beta42*120*120));
  /*At t = 200*/
  estimate"P(0)4"exp(int1+beta11*200+beta21*3.5392+beta31*200*3.5392+beta41*200*200)
  /(1+exp(int1+beta11*200+beta21*3.5392+beta31*200*3.5392+beta41*200*200));
  estimate"P(<=1)"exp(int2+beta12*200+beta21*3.5392+beta31*200*3.5392+beta42*200*200)
  /(1+ exp(int2+beta12*200+beta21*3.5392+beta31*200*3.5392+beta42*200*200));
  estimate"P(2)"1-exp(int2+beta12*200+beta21*3.5392+beta31*200*3.5392+beta42*200*200)
  /(1+ exp(int2+beta12*200+beta21*3.5392+beta31*200*3.5392+beta42*200*200));
run; /*-2LogL = 321.5. -2DeltaL = 329.4-321.5 = 7.9
p-value = 0.0195*/
```

```
    /*Make prediction curve*/
proc nlmixed data = phn_com qpoints = 20;
  parms beta11 = -0.09716 beta21 = -0.6936 beta31 = 0.01573 beta41 = 0.000207
        int1 = 3.7085  int2 = 4.2871 d = 2
        beta12 = -0.09716 beta42 = 0.000207;
  eta1 = beta11*time + beta21*LOGAB + beta31*time*LOGAB + beta41*time2 + b1;
  eta2 = beta12*time + beta21*LOGAB + beta31*time*LOGAB + beta42*time2 + b1;
  /*random intercept and slope*/
  if phn = 0 then z = 1/(exp(-eta1 - int1) + 1);
  else if phn = 1 then z = 1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1);
  else if phn = 2 then z = 1-1/(exp(-eta2 - int2) + 1);
  ll = log(z);

  model phn ~ general(ll);
  random b1 ~ normal(0, d*d) subject = id out = RE;
  predict  1/(exp(-eta1 - int1) + 1) out = prediction0;
  /*Probability of having PHN = 0*/
  predict  1/(exp(-eta2 - int2) + 1) - 1/(exp(-eta1 - int1) + 1) out = prediction1;
  predict  1-1/(exp(-eta2 - int2) + 1) out = prediction2;
      /*Probability of having PHN = 2*/
  predict 1/(exp(-eta2-int2)+1) out = prediction01;
      /*Probability of having PHN <= 1*/
run;
                     /*VISUALIZE THE DATA*/
  /*Predicted curve and Observed curve for probability PHN = 2*/
data phn_com_PHN; set phn_com;
if PHN = 2 then PHN2 = 1; else if PHN = 1 | PHN = 0 then PHN2 = 0; else PHN2 = PHN;
if PHN = 0 then PHN0 = 1; else if PHN = 1 | PHN = 2 then PHN0 = 0; else PHN0 = PHN;
run;

/*LOESS: Probability of having PHN = 2 : Observed Data*/
proc loess data = phn_com_PHN;
    ods output scoreresults = out2obs; model phn2 = time logab /
        scale = sd(0.1) details(ModelSummary OutputStatistics);
score data = phn_com_PHN;
run;  /*Smoothing parameter = 0.48712*/
proc sort data = out2obs; by time; run;
proc means data = out2obs; var p_phn2; run;
    /*LOESS: Probability of having PHN = 2: Predicted Data*/
proc loess data = prediction2;
    ods output scoreresults = out2; model pred = time logab / scale = sd(0.1);
score data = prediction2; run; /*Smoothing parameter = 0.44421*/
proc sort data = out2; by time; run;
proc means data = out2; var p_pred; run;
```

```
goptions reset = all ;
data combine2; set out2 out2obs; keep time pred p_pred phn2 p_phn2 logab; run;
legend1 label=none
        position=(bottom center outside)
        ;
proc gplot data = combine2;
    plot phn2*time = 1 p_phn2*time = 2
    pred*time = 3 p_pred*time = 4 /
    overlay haxis = axis1 vaxis = axis2 legend = legend1;
symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
symbol3 c = green v = dot h = 0.4 mode = include;
    symbol4 c = blue i = join w = 2 mode = include;
    axis1 label = (h=2 "Time since entry") value=(h=1.5)  minor=none ;
axis2 label = (h=2 A=90 "Probability of having PHN = 2")
        value=(h=1.5)  minor = none;
run;

    /*Use the same smooth parameter 0.465*/
proc loess data = phn_com_PHN; ods output scoreresults = out2obs_smooth;
    model phn2 = time logab /scale = sd(0.1) smooth = 0.465;
    score data = phn_com_PHN; run;
proc sort data = out2obs_smooth; by time; run;

proc loess data = prediction2; ods output scoreresults = out2_smooth;
    model pred = time logab/ scale = sd(0.1) smooth = 0.465;
    score data = prediction2; run;
proc sort data = out2_smooth; by time; run;

data combine2_smooth; set out2_smooth out2obs_smooth;
keep time pred p_pred phn2 p_phn2 logab; run;
    /*Plot probability over time*/
proc gplot data = combine2_smooth;
    plot phn2*time = 1 p_phn2*time = 2
    pred*time = 3 p_pred*time = 4 /
    overlay haxis = axis1 vaxis = axis2 legend = legend1;
symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
symbol3 c = green v = dot h = 0.4 mode = include;
    symbol4 c = blue i = join w = 2 mode = include;
    axis1 label = (h=2 "Time since entry") value=(h=1.5)  minor=none ;
axis2 label = (h=2 A=90 "Probability of having PHN = 2")
    value=(h=1.5)  minor = none;
run;
```

```
    /*Plot probability over logab*/
proc sort data = combine2_smooth; by logab; run;
proc gplot data = combine2_smooth;
    plot phn2*logab = 1 p_phn2*logab = 2
    pred*logab = 3 p_pred*logab = 4 /
    overlay haxis = axis1 vaxis = axis2 legend = legend1;
symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
symbol3 c = green v = dot h = 0.4 mode = include;
    symbol4 c = blue i = join w = 2 mode = include;
    axis1 label = (h=2 "LOGAB over time") value=(h=1.5)  minor=none ;
axis2 label = (h=2 A=90 "Probability of having PHN = 2")
    value=(h=1.5)  minor = none;
run;
```

```
              /*Predicted curve and Observed curve for probability PHN = 0*/
              /*LOESS: Probability of having PHN = 0 : Observed Data*/
proc loess data = phn_com_PHN;
    ods output scoreresults = out0obs; model phn0 = time logab;
score data = phn_com_PHN;
run; /Smoothing parameter = 0.50858;
proc sort data = out0obs; by time; run;
proc means data = out0obs; var p_phn0; run;
 /*LOESS: Probability of having PHN = 0: Predicted Data*/
proc loess data = prediction0;
    ods output scoreresults = out0; model pred = time logab;
score data = prediction0; run; /Smoothing parameter = 0.62446;
proc sort data = out0; by time; run;
proc means data = out0; var p_pred; run;

goptions reset = all ;
data combine0; set out0 out0obs; keep time pred p_pred phn0 p_phn0 logab; run;
legend1 label=none
        position=(bottom center outside)
        ;
proc gplot data = combine0;
    plot phn0*time = 1 p_phn0*time = 2
```

```
    pred*time = 3 p_pred*time = 4 /
    overlay haxis = axis1 vaxis = axis2 legend = legend1;
symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
symbol3 c = green v = dot h = 0.4 mode = include;
    symbol4 c = blue i = join w = 2 mode = include;
    axis1 label = (h=2 "Time since entry") value=(h=1.5)  minor=none ;
axis2 label = (h=2 A=90 "Probability of having PHN = 0")
    value=(h=1.5)  minor = none;
run;

     /*Use the same smooth parameter*/
proc loess data = phn_com_PHN;
    ods output scoreresults = out0obs_smooth; model phn0 = time logab/
    scale = sd(0.1) smooth = 0.5663;
score data = phn_com_PHN;
run;
proc sort data = out0obs_smooth; by time; run;

/*LOESS: Probability of having PHN = 2: Predicted Data*/
proc loess data = prediction0;
    ods output scoreresults = out0_smooth; model pred = time logab/
    scale = sd(0.1) smooth = 0.5663;
score data = prediction0; run;
proc sort data = out0_smooth; by time; run;

goptions reset = all ;
data combine0_smooth; set out0_smooth out0obs_smooth;
keep time pred p_pred phn0 p_phn0 logab; run;
legend1 label=none
        position=(bottom center outside)
        ;
 /*Plot probability over time*/
proc gplot data = combine0_smooth;
    plot phn0*time = 1 p_phn0*time = 2
    pred*time = 3 p_pred*time = 4 /
    overlay haxis = axis1 vaxis = axis2 legend = legend1;
symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
symbol3 c = green v = dot h = 0.4 mode = include;
    symbol4 c = blue i = join w = 2 mode = include;
    axis1 label = (h=2 "Time since entry") value=(h=1.5)  minor=none ;
axis2 label = (h=2 A=90 "Probability of having PHN = 0")
    value=(h=1.5)  minor = none;
run;
```

```
         /*Plot probability over logab*/
proc sort data = combine0_smooth; by logab; run;
proc gplot data = combine0_smooth;
    plot phn0*logab = 1 p_phn0*logab = 2
    pred*logab = 3 p_pred*logab = 4 /
    overlay haxis = axis1 vaxis = axis2 legend = legend1;
symbol1 c = red v = dot h = 0.4 mode = include;
    symbol2 c = black i = join w = 2 mode = include;
symbol3 c = green v = dot h = 0.4 mode = include;
    symbol4 c = blue i = join w = 2 mode = include;
    axis1 label = (h=2 "LOGAB over time") value=(h=1.5)  minor=none ;
axis2 label = (h=2 A=90 "Probability of having PHN = 0")
    value=(h=1.5)  minor = none;
run;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Longitudinal modeling of antibody dynamics during herpes zoster infection

Richting: **Master of Statistics-Epidemiology & Public Health Methodology**
Jaar: **2015**

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,

Tran, Mai Phuong Thao

Datum: **1/09/2015**