

Non-linear Mixed Models in a Dose Response Modelling

Madona Yunita Wijaya

Department of Mathematics, Faculty of Sciences and Technology
Universitas Islam Negeri Syarif Hidayatullah Jakarta
Email: madona@uinjkt.ac.id

Abstract

Study designs in which an outcome is measured more than once from time to time result in longitudinal data. Most of the methodological works have been done in the setting of linear and generalized linear models, where some amount of linearity is retained. However, this still be considered a limiting factor and non-linear models is another option offering its flexibility. Non-linear model involves complexity of non-linear dependence on parameters than that in the generalized linear class. It has been utilized in many situations such as modeling of growth curves and dose-response modeling. The latter modeling will be the main interest in this study to construct a dose-response relationship, as a function of time to IBD (inflammatory bowel disease) dataset. The dataset comes from a clinical trial with 291 subjects measured during a 7 week period. Both linear and non-linear models are considered. A dose time response model with generalized diffusion function is utilized for the non-linear models. The fit of non-linear models are found to be more flexible than linear models hence able to capture more variability present in the data.

Keywords: IBD; longitudinal; linear mixed model; non-linear mixed model.

Abstrak

Desain studi dimana hasil diukur berulang kali dari waktu ke waktu menghasilkan data longitudinal. Sebagian besar metodologi yang digunakan untuk menganalisis data longitudinal adalah model linear dan model linear umum (generalized linear model) dimana sejumlah linearitas tertentu dipertahankan. Asumsi linearitas ini masih dipandang memiliki keterbatasan dan model non-linear adalah pilihan metode lainnya yang menawarkan fleksibilitas. Model non-linear telah digunakan di berbagai macam situasi seperti model kurva pertumbuhan, model farmakokinetika, dan farmakodinamika, dan model respon-dosis. Model respon-dosis akan menjadi fokus dalam penelitian ini untuk membangun hubungan dosis-respon sebagai fungsi waktu dari data IBD dengan menggunakan model linear dan non-linear. Hasil penelitian menunjukkan bahwa model non-linear lebih fleksibel daripada model linear sehingga mampu menangkap lebih banyak variabilitas yang ada di dalam data.

Kata kunci: IBD; longitudinal; model linear; model non-linear.

1. INTRODUCTION

The generalized linear models (GLM) is an extension of the classical linear normal [1]. In LM and GLM settings, a certain amount of linearity is preserved; the former having a linear relationship between the mean response and the linear predictor, and the latter, being linear at the predictor. These models are very flexible and very popular among researchers where they have received much attention in the past few decades as compared to non-linear models. However, there has been increased need for advance modelling technique including non-linear modelling methods. Some of notable reasons

are: first, modern statistics are confronted with complex data structures which become increasingly available with modern computing power. Secondly, within GLM framework, of which the LM is a special case with a normal distribution and the identity link, one is limited to the choice of distributions only from the exponential family. Hence non-linear models provide a more extensive option. Thirdly, with non-linear models, one captures more shapes with few parameters since few parameters generate a vast majority of shapes. Lastly, linearity is in most cases, just an approximation which may be less meaningful in some situations. In these kind of situations involve phenomena of growth over sufficiently extended periods, particularly when the observational period includes both growth spurts and asymptotic behavior of growth toward maturation. Dose-response modelling, pharmacokinetic, and pharmacodynamic applications often demand non-linear models as well [2]. The non-linearity is not restricted to the fixed effects but sometimes includes the random effects. A flexible framework to accommodate various deviation from the general linear mixed model is the non-linear mixed model of which the generalized linear mixed model is a special case with non-linearity in link function but linear in the predictors.

In this study, we will apply both linear and non-linear models to inflammatory bowel disease (IBD) dataset that was obtained from a longitudinal observational study. IBD commonly refers to ulcerative colitis (UC) and Crohn disease (CD), which are chronic inflammatory diseases of the GI tract of unknown etiology [3] [4] [5]. The small intestine is usually affected by the Crohn disease, but the affect can be more widespread through the system of gastrointestinal. On the other hand, Ulcerative Colitis is confined to the rectum and colom. Most of the time, it involves the rectum alone. Inflammation and ulceration of the lining of the bowel cause urgent diarrhea which can be very frequent. Both of these conditions need treatments of anti-inflammatory drug, but the choice of drugs are rather different between the two. Steroids are often needed either locally in the rectum or orally to treat acute attacks. Frequently relapsing cases will need immunosuppressive treatments such as azathioprine.

The aim of this study is to construct a dose-response relationship, as a function of time to IBD dataset as described in Section 2. The method that will be used is discussed in Section 3 including linear and non-linear mixed models. The findings are explained in Section 4 and the discussion and conclusion in Section 5.

2. DATA DESCRIPTION

The dataset comes from a clinical trial with 291 subjects, divided over four treatments: placebo (0), 1000mg (1), 2000mg (2), and 4000mg (3). Patients are measured during a 7 week period. The outcome of interest is an IBD activity score (*IBDSC*). In total, there are 4 variables in the dataset, i.e. Week (measured at weeks 1 through 7), Treat (Treatment indicator: 0, 1, 2, 3), *IBDSC* (IBD score), and *IBDSC0* (IBD score at baseline). Note that not all of the patients were followed up until the end of the study resulting in missing measurements for some time points as can be seen in Table 1. It is also observed that 64.9% of the profiles are complete, while 7.6% exhibit dropouts and the remaining 27.5% have intermittent missing values.

Table 1. Summary of observations taken at each time point

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
#Obs.	291	279	265	247	231	221	279
Percentage	100%	96%	91%	85%	79%	76%	96%

3. METHOD

3.1. Linear Mixed Model

In practice we are often confronted with unbalanced data, therefore multivariate regression techniques are no longer applicable to analyze many longitudinal data sets [6]. This leads to a two-stage model where in the first stage the evolution of response variables is modeled for each patient or subject using a linear regression model. In second stage, the subject-specific regression coefficients obtained in stage one are used as response variables and is fitted with other known covariates [6]. The first stage assumes that \mathbf{Y}_i follows the linear regression model:

$$\mathbf{Y}_i = \mathbf{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \tag{1}$$

where \mathbf{Z}_i is $(n_i \times q)$ matrix of known covariates. It models how the response behaves over time for the i^{th} subject. The $\boldsymbol{\beta}_i$ is a q -dimensional vector of unknown subject specific regression coefficients, and $\boldsymbol{\varepsilon}_i$ is a vector of residual components ε_{ij} , $j=1,2,\dots,n_i$. In a second stage, a multivariate regression model of the form:

$$\boldsymbol{\beta}_i = \mathbf{K}_i \boldsymbol{\beta} + \mathbf{b}_i \tag{2}$$

is employed to explain the observed variability between the subjects, with respect to their subject-specific regression coefficients $\boldsymbol{\beta}_i$, \mathbf{K}_i is a $(q \times p)$ matrix of known covariates, and $\boldsymbol{\beta}$ is a p -dimensional vector of unknown regression parameters. In addition, the \mathbf{b}_i are assumed to be independent and follow a q -dimensional normal distribution with mean vector zero and general covariance matrix D [6].

Due to the extra variability and loss of information experienced in the two-stage analysis, the random-effects models that combines the two steps is applied and is defined as follows [6].

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i \tag{3}$$

where $\mathbf{X}_i = \mathbf{Z}_i \mathbf{K}_i$ is a matrix of known covariates and where all other components are as defined earlier with:

$$\begin{cases} \mathbf{b}_i \sim N(\mathbf{0}, D), \\ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma_i), \\ \mathbf{b}_i \text{ and } \boldsymbol{\varepsilon}_i \text{ are independent.} \end{cases} \tag{4}$$

The selection of a random-effects model is done by selecting the preliminary structures for the mean, for random-effects, and for the residual covariance. In order to find the appropriate random effect, we need to conduct a significance test of the highest order random effect first in a hierarchical way. Therefore, the Restricted Maximum Likelihood (REML) test will be applied to test whether the random effects are needed since this will be based on the same mean structure. The classical likelihood-based inference to test for the need of random effects cannot be applied because the corresponding hypothesis is on the boundary of the parameter space. Thus, the asymptotic null distribution of the likelihood ratio test statistic is a mixture of chi-square with equal weights 0.5. The Maximum Likelihood (ML) approach will be used in the mean structure reduction since the REML approach breaks down. This is due to the fact that different mean structures are compared, which will yield incomparable results in case the REML is used.

3.2. Non-linear Mixed Model

In several situations, statistical models where the mean is not defined as a function of a linear predictor are needed. These are called non-linear mixed models that can take various forms, but the most common ones involve a conditional distribution of \mathbf{Y}_{ij} given \mathbf{b}_i belongs to the exponential family, encompassing the outcomes that come from normal and non-normal distribution. The mean of the model can be generally expressed as $E(Y_{ij} | b_i) = h(x_{ij}, \beta, z_{ij}, b_i)$. Similar to generalized linear mixed models (GLMM), assuming the random effects to follow normal distribution with mean $\mathbf{0}$ and covariance matrix \mathbf{D} is often considered, even though assuming from other distributions are possible in principle. The same methods can be utilized to estimate the parameter as were developed for GLMM [7].

Based on statistical literature review, the following dose time-response model (generalized diffusion function) is proposed to fit to the IBD dataset [8] [9] [10]:

$$Y_{ij} = A_0 + r_{1i} + \frac{(B_0 + r_{2i} + B_1 \times \text{dose}_i) \times \text{week}_{ij}}{1 + \exp(C_1 \times \text{dose}_i) \times \text{week}_{ij}^{D_0}} + \varepsilon_{ij}, \quad (5)$$

where $r_{1i} \sim N(0, \sigma_{r_1}^2)$, $r_{2i} \sim N(0, \sigma_{r_2}^2)$, and $\varepsilon_{ij} \sim N(0, \sigma^2)$. The term $(B_0 + B_1 \times \text{dose}_i) \exp(C_1 \times \text{dose}_i)$ is the dose-dependent asymptote in time and $\exp(C \times \text{dose}_i)$ is the dose-dependent half life time. As such it prescribes a trajectory of persistent dose-effects, i.e. dose effects sustain at a constant level and do not show any recovery from the dose effects within the experimental time course [11]. Due to computational difficulty, only two subject specific or random effects are considered, i.e. attaching r_1 and r_2 to A_0 and B_0 , respectively, and they are assumed to be independent.

4. RESULTS

4.1. Exploratory Data Analysis (EDA)

The individual profiles from 40 randomly selected patients for each treatment groups are displayed in Figure 1. It can be observed that there are much variability between and within patients suggesting the need of random intercept and random slope in the model.

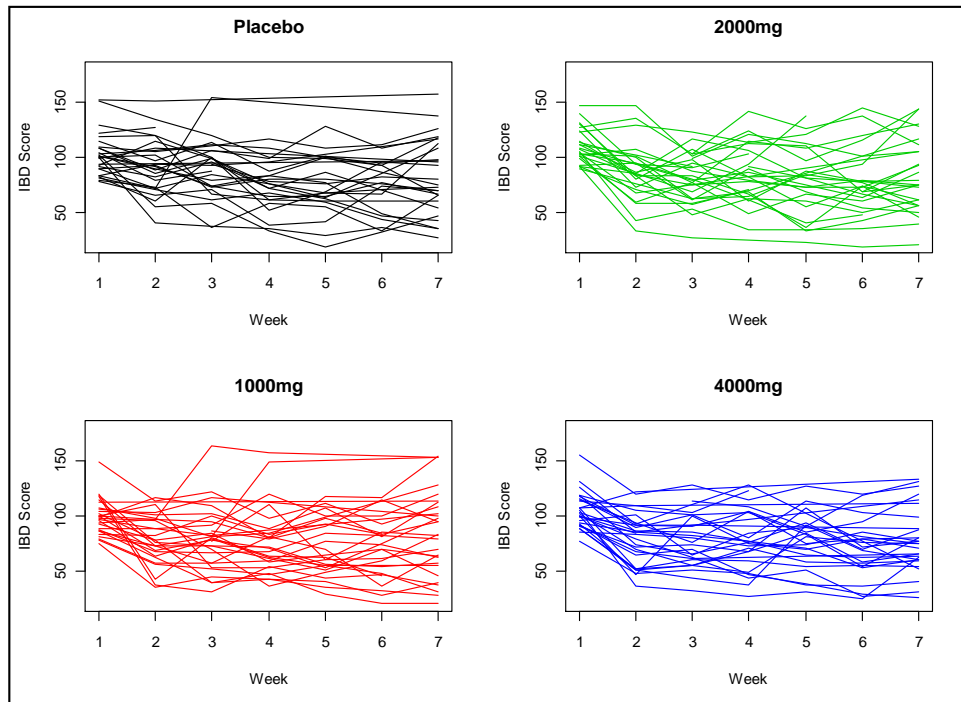


Figure 1. Individual profiles per treatment groups.

It is also confirmed by the non-constant variance structure across the time (Figure 2, right panel). The variances seem to increase with time and this could be due to attrition. This suggests that caution should be used with incomplete data. In addition, mean structure for each treatment indicates the need of non-linear function to describe the evolution. They seem to follow a quadratic evolution over time (Figure 2, left panel) and this can be considered to construct our initial model for linear mixed model in Section 4.2.

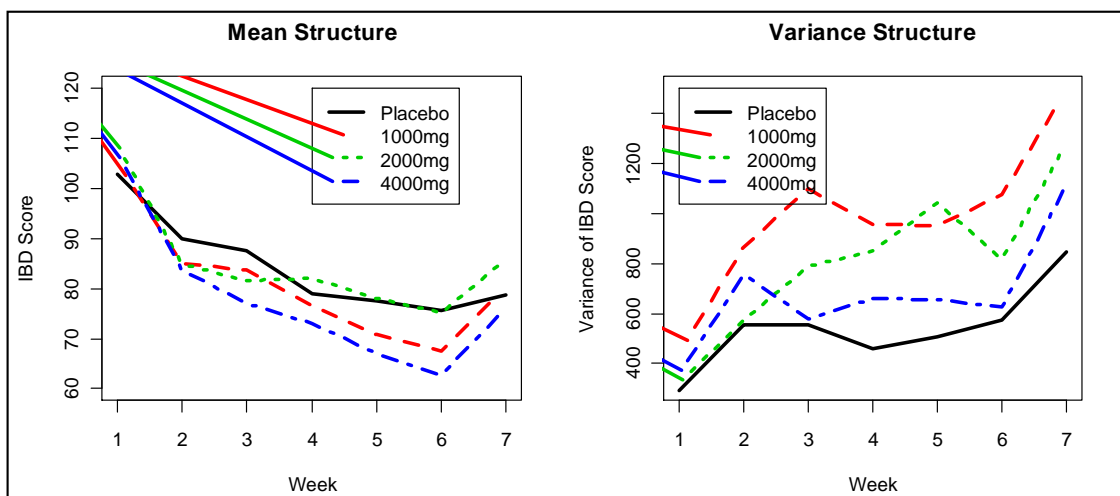


Figure 2. Mean and variance structures for the different treatment groups

4.2. Linear Mixed Model

The adequacy of the first stage model based on graphical exploration in Section 4.1 was explored by testing for the need of a model extension. There is strong evidence in favor using the quartic first stage model results in $F_{meta} = 1.105$ which is not significant (p -value = 0.2402) when compared to an F -distribution with 224 and 189 degrees of freedoms. In the second stage model, the subject-specific intercepts and time effects are related to the dose levels as a continuous rather than as a categorical variable assuming the placebo group received dose level = 0. This was considered since no substantial differences obtained from the model by treating the dose as continuous or categorical. Thus by taking continuous dose level, we do not lose information. Further efficiency is achieved by estimating less parameter.

The preliminary linear mixed effects model is suggested by combining the first and second stage models with random intercept and random slopes with unstructured working assumption. Likelihood Ratio Test (LRT) statistics with the correct null distribution was performed to test the need of random effects. The model is then reduced to a more parsimonious model by deleting insignificant terms in a hierarchical way. The results for the final model based on the observed data (direct likelihood) are shown in Table 2 assuming the missingness in the data is at random. The final model was also fitted after generating 5 multiple imputations. Both models yielded similar parameter estimates and led to the same inference.

4.3. Non-linear Mixed Model

An alternative to linear mixed model to incorporate a more flexible function that can approximate the observed mean profile is called non-linear mixed model. Parameter estimates and standard errors obtained from fitting model (5) are summarized in Table 3. It can be observed that all parameters are significant. The model was tried to simplify by removing the random η_2 effect but it was not possible since the likelihood ratio equals 151 on 0 and 1 degree of freedom ($p < 0.0001$). It can be seen that the fitted model is non-linear in the fixed-effects parameters and linear in the random effects. Thus, the calculated marginal mean over the distribution of random-effects is simplified.

Table 2. The estimated parameter (standard error) for the linear mixed model

Parameter	Direct Likelihood		Multiple Imputation	
	Estimate (s.e.)	P-value	Estimate (s.e.)	P-value
Intercept	150.150 (5.394)	<.0001	149.904 (5.899)	<.0001
Dose	2.863 (1.132)	0.012	2.757 (1.172)	0.0193
Week	-68.040 (7.665)	<.0001	-67.333 (8.704)	<.0001
Week ²	26.152 (3.500)	<.0001	25.682 (4.192)	<.0001
Week ³	-4.309 (0.638)	<.0001	-4.204 (0.794)	<.0001
Week ⁴	0.252 (0.040)	<.0001	0.245 (0.051)	0.0001
Dose × week	-2.177 (0.721)	0.003	-2.143 (0.789)	0.0087
Dose × week ²	0.233 (0.083)	0.005	0.231 (0.091)	0.0145

**dosage unit is g instead of mg*

Table 3. The estimated parameter for the non-linear mixed model

Parameter	Estimate	Std. Error	P-value
A ₀	141.620	6.8184	<.0001
B ₀	-77.617	12.671	<.0001
B ₁	3.907	1.9592	0.0471
C ₁	-0.079	0.0278	0.0046

D_0	1.045	0.0355	<.0001
$\text{Var}(r_1)$	39.650	16.293	0.0156
$\text{Var}(r_2)$	789.67	111.65	<.0001
σ	16.450	0.3079	<.0001

4.4. Model Comparison

The marginal profiles obtained from both linear and non-linear mixed models are depicted in Figure 3 for each dose level/treatment group.

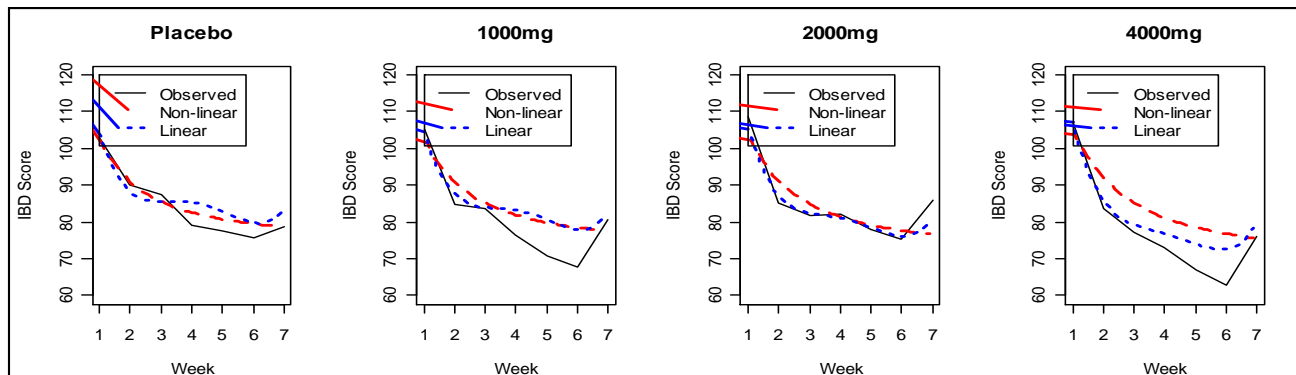


Figure 3. Predicted and observed average profiles per treatment group

The linear mixed model shows a good fit at the beginning and at the end of the study and seems to overfit in the middle of the study for the four treatment groups, especially for 1000mg and 4000mg. The same picture is delineated from fitting the non-linear mixed model. In addition, less parameter is needed to be estimated under nonlinear mixed model compare to linear mixed model. Thus, we achieve higher efficiency by fitting non-linear mixed model to come up with excellent results. The four different dose levels show similar trend, i.e. higher IBD score in week 1 and the score decreases as time increases. Higher dose level implies lower IBD score.

5. DISCUSSION AND CONCLUSION

In this paper, it is of interest to explore the dose-response relationship as a function of time. When conventional linear models may be insufficient, non-linear models can be used as an alternative to describe the evolution of the profiles in a longitudinal setting. Linear and non-linear mixed models are used to fit to the IBD dataset and both approaches seem to fit very well although they seem to overfit in the middle of the study as seen in Figure 3. This could be due to intermittent missing value and it was observed that 27.5% of the patients have non-monotone pattern. However, linear mixed model might be inadequate in this case since taking a higher order polynomial can impose multiple peaks or valley on the dose-time-response, and is difficult to interpret^[8]. Meanwhile, non-linear mixed model offers efficiency; only need less parameter than linear mixed model. The user also has the flexibility to define the nonlinear model that applies to a particular data set or response although it is not an easy task. But it will be very handy if we have a theoretical background.

REFERENCES

[1] J.A. Nelder and R.W.M. Wedderburn, "Generalized linear models.," *Journal of the Royal Statistical Society : Series (A) General*, vol. 153, no. 3, pp. 370-384, 1972.

- [2] G. Molenberghs and G. Verbeke, "Models for Discrete Longitudinal Data," New York: Springer, 2005.
- [3] N.J. Talley, M.T. Abreu, J.P. Achkar, C.N. Bernstein, M.C. Dubinsky, S.B. Hanauer, S.V. Kane, W.J. Sandborn, T.A. Ullman, and P. Moayyedis, "An Evidence-Based Systematic Review On Medical Therapies for Inflammatory Bowel Disease," *Am J Gastroenterol*, vol. 106, pp. S2-S25, April 2011.
- [4] C. Abraham and J.H. Cho, "Inflammatory Bowel Disease," *The New England Journal of Medicine*, vol. 361, no. 21, pp. 2066-2078, November 2009.
- [5] R.C. Langan, P.B. Gotsch, M.A. Krafczyk, and D.D. Skillinge, "Ulcerative Colitis: Diagnosis and Treatment," *American Family Physician*, vol. 76, no. 9, pp. 1323-1330, November 2007.
- [6] G. Verbeke and G. Molenberghs, "Linear Mixed Model for Longitudinal Data.," , New York: Springer, 2000.
- [7] J. Serroyen, G. Verbeke, G. Molenberghs, and M. Davidian, "Non-Linear Models for Longitudinal Data," *American Statistician*, vol. 63, no. 4, pp. 378-388, 2009.
- [8] S. Woodruff, "The Use of Mixed-Effects Dose-Response Modeling for Neurotoxicity Risk Assessment," University of South Florida, Master's Thesis, 2001.
- [9] Y. Zhu, "Neurobehavioral Toxicity Risk Assessment: Dose-Time-Response Modeling and Benchmark Dose Estimation," National Center for Environmental Assessment, US EPA, 2001.
- [10] Y. Zhu, "Technical Guidance Document for Dose-Response Modeling and Benchmark Dose Estimation of Neurobehavioral Toxicity Screening Data," National Center for Environmental Assessment, US EPA, 2003.
- [11] Y. Zhu, M.R. Wessel, T. Liu, and V.C. Moser, "Analysis of Neurobehavioral Screening data: Dose-Time-Response Modeling of Continuous Outcome," *Regulatory Toxicology Pharmacology*, vol. 41, no. 3, pp. 240-255, 2005.