IS4250 HEALTHCARE ANALYTICS

Project Report

Semester 2, AY 2015/16

Prepared

for

Prof. Chan Wai Mun, Raymond

Group 02

Lin Baoyu A0105715M

Zhang Yaowen A0105745H

Table of Contents

1. Contributions

Nowadays, biologics such as monoclonal antibodies (mAbs) are increasingly used for the treatment of many chronic diseases, including cancer (Hoos et al., 2007) and auto-immune diseases (Winthrop et al., 2015). Numerous cases have shown that there are significant improvement in the treatment outcomes by using biologics. Currently, more than 30 mAbs have been approved by the administrative government agencies. Hundreds of clinical trials with mAbs have been conducted, and more effort will be spent on the mABs studies and clinical trails in the future.

Treatment plans for mAbs are much more complicated than that for traditional medications. MAbs are injected at time long intervals, typically weeks or months, which are much longer than that of conventional medicines. Therefore, unlike the dose–response models for traditional medications, dose-time-response models are required to describe and predict the drug effect for complex mAbs treatment plan development.

To identify an adequate regimen for the investigated mAb, multiple clinical trials with different doses and injection times are conducted to compare different regimens. However, this kind of clinical trails can hardly to be complete and efficient, because the investigated regimens can cover only a small proportion of all feasible combinations of doses and time-intervals.

Hence, reliable predictions about the response for untested regimens will be of great help in the design of the following clinical trails. Based on predictions for several considered alternative regimens, the most promising regimen can then be highlighted and evaluated in subsequent clinical trials. Therefore, an accurate dose-time-response model will significantly improve the efficiency of the traditional clinical trails.

In *Analysis of clinical trials with biologics using dose-time-response models*, Lange and Schmidli (2015) constructed a dose-time-response model, using semi-mechanistic nonlinear regression models, to describe and predict the time-changing response for complex dosing regimens. This paper provided a guideline for future dose-time-response relationship modeling studies.

2. Methodology

(1) To build a simple and basic dose-time-response model based on pharmacokinetic–pharmacodynamic (PKPD) models, where the PK component in the unobserved effect compartment is treated as a latent variable

(2) To extend the basic dose-time-response model by including more factors that will potentially affect the response

(3) To estimate and predict the parameters using Maximum-likelihood (ML) and Bayesian methods

(4) To evaluate frequentist operating characteristics of the methods by using simulations

(5) To analyze clinical trails based on the predictions of the dose-time-response model

3. The Dose-Time-Response Model

The basic dose-time-response model relies on simplified mechanistic concepts. The latent kinetic component C(t) describe the drug concentration in the unobserved effect compartment. MAbs are usually administered as subcutaneous injections, as illustrated by Figure 1.
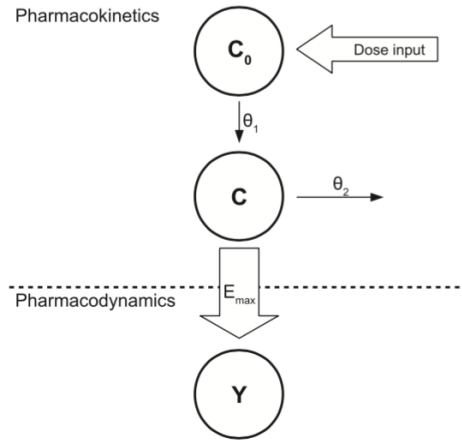
Figure 1. Schematic illustration of the pharmacokinetics of a biologic injected in a subcutaneous skin depot C0. The resulting concentration in the unobserved effect compartment C is then directly linked to the response Y via the E_{max} model.

C_0 represents the initial concentration when the drug is injected. The biologic is then transferred to the unobserved effect compartment C at an absorption rate θ_1. At the same time, the drug effect is disappear at a rate θ_2. This process can be concluded with the following differential equation

$$\frac{dC_0(t)}{dt} = -\theta_1 C_0(t)$$

$$\frac{dC(t)}{dt} = -\theta_2 C(t) + \theta_1 C_0(t).$$

2

Therefore, for a single dose D given at time 0, the concentration in the effect compartment over time is then

$$C(t) = \frac{D\,\theta_1}{\theta_1 - \theta_2}\left(e^{-\theta_2 t} - e^{-\theta_1 t}\right)$$

Base on direct-response E_{max}-model, we have following equation:

$$g(C(t)) = \theta_5 + \frac{\theta_4 \cdot C(t)}{\theta_3 + C(t)}.$$

Here, θ_3 is the concentration which produces a 50% maximum response. θ_4 is the maximum possible response, θ_5 is the expected response of baseline.

After simplify the equation above, the basic model is constructed. If patient i get an injection with dose Di and a continuous response y_{ij} at time t_j will be:

$$y_{ij} = \theta_5 + \frac{\theta_4 D_i\left(e^{-\theta_2 t_j} - e^{-\theta_1 t_j}\right)}{\theta_3\left(1 - \theta_2/\theta_1\right) + D_i\left(e^{-\theta_2 t_j} - e^{-\theta_1 t_j}\right)} + \epsilon_{ij}$$

The paper also discusses six important extensions of the basic model considering different factors of placebo response, baseline covariates, multiple doses, route of administration, residual error, and indirect-response models & turnover models.

To use ML methods to derive the parameters of the model, numerical optimization procedures are necessary. As the likelihood function may have several local maxima, a good initial guess of parameter value is critical. To derive confidence and prediction intervals is based on asymptotic theory and linear approximations. Hence, the validity of intervals is questionable for small sample size. Moreover, if the true parameter is close to boundary of parameter space, the asymptotic ML theory is not applicable. For dose-time-response models, it is common that the parameters are actually on the boundary.

For Bayesian inference, incorporate the knowledge of previous clinical trial by using informative priors could be useful in estimating parameters. However, the posterior distribution calculation requires multidimensional integration. By using Markov chain Monte Carlo (MCMC) methods, we can get a distribution of sample posterior.

4. Simulation to Evaluate Operating Characteristics

In the clinical trial simulation, 1000 datasets are generated using a multivariate normal distribution. Then the operating characteristics are evaluated by statistical methods based on the data obtained. E_{max} model is used to assess the expected value using the true values as the parameters, and the results are shown in Table 1.

		Likelihood analysis		Bayesian analysis	
Parameter	True value	Median of maximum likelihood-estimate	95% CI coverage (in %) *	Median of posterior median	95% PI coverage (in %) *
θ_1	0.07	0.067	90.7	0.087	94.8
θ_2	0.03	0.032	91.2	0.028	97.2
θ_3	10.00	9.67	97.4	13.23	92.1
θ_4	−80.00	−80.7	99.7	−82.16	92.5
θ_5	100.00	100.0	91.9	98.92	93.6
Time					
0	100.00	100.0	91.9	98.92	93.6
10	28.38	27.9	99.8	26.80	95.3
50	30.32	30.5	99.9	31.32	94.6
100	26.18	25.4	100.0	24.44	93.5
200	57.13	57.0	98.4	57.49	94.3
300	95.61	94.8	94.6	92.21	90.5

Table I. Summary results of the clinical trial simulations (CI=confidence interval, PI=probability interval).

* Monte Carlo standard error: 0.7%

Both the ML and the Bayesian methods show that their point estimates are near to the true parameter values, where ML estimates are closer to the true values on average. Furthermore, the confidence and probability interval coverage are acceptable here, with the Bayesian probability interval coverage slightly better than ML confidence interval coverage.

5. Clinical Trails

Dose-time-response models are used to analyze a clinical trail where the effects of canakinumab subcutaneous injection were examined. Active control was used in this 24-week double-blinded dose-ranging study.

The 400 participants were randomly divided into seven groups, where five of them received single doses (either 25 mg, 50 mg, 100 mg, 200 mg, or 300 mg at day 1), one group received multiple doses (50 mg at day 1, 50 mg at week 4, 25 mg at week 8, and 25 mg at week 12), and the last group was the placebo. The logarithm of C-reactive protein (CRP) levels was used as the response in this study, where CRP is a biomarker signaling the severeness of the disease.

The data collected were used to investigate: (1) The fitness of the dose-time-response model in the clinical trail; (2) The accuracy of the dose-time-response model predictions of the untested treatment regimens response.

Ideally, the model should describe the data well for each single-dose group, and also predict the actual observed response of the treatment group with multiple injections with a reasonable level of accuracy.

Table II shows the evaluation using the Bayesian analysis for the five single-dose treatment groups. Here, θ_6 is added to the basic model to measure the effect of the baseline CRP levels (X_0).

Table II. Bayesian analysis for data from the five single-dose treatment groups.		
	Posterior median	95% probability interval
θ_1	0.177	0.077 — 0.293
θ_2	0.022	0.016 — 0.031
θ_3	8.07	3.44 — 15.19
θ_4	−1.41	−1.82 — −1.06
θ_5	0.934	0.65 — 1.29
θ_6	0.431	0.38 — 0.53

To assess the fitness of the dose-time-response model, observed values and the fitted Bayesian model are compared in Figure 2. The last plot displays all posterior median curves relative to their baseline value. The dose-time-response model describes the single-dose data well with $R^2 = 0.82$.
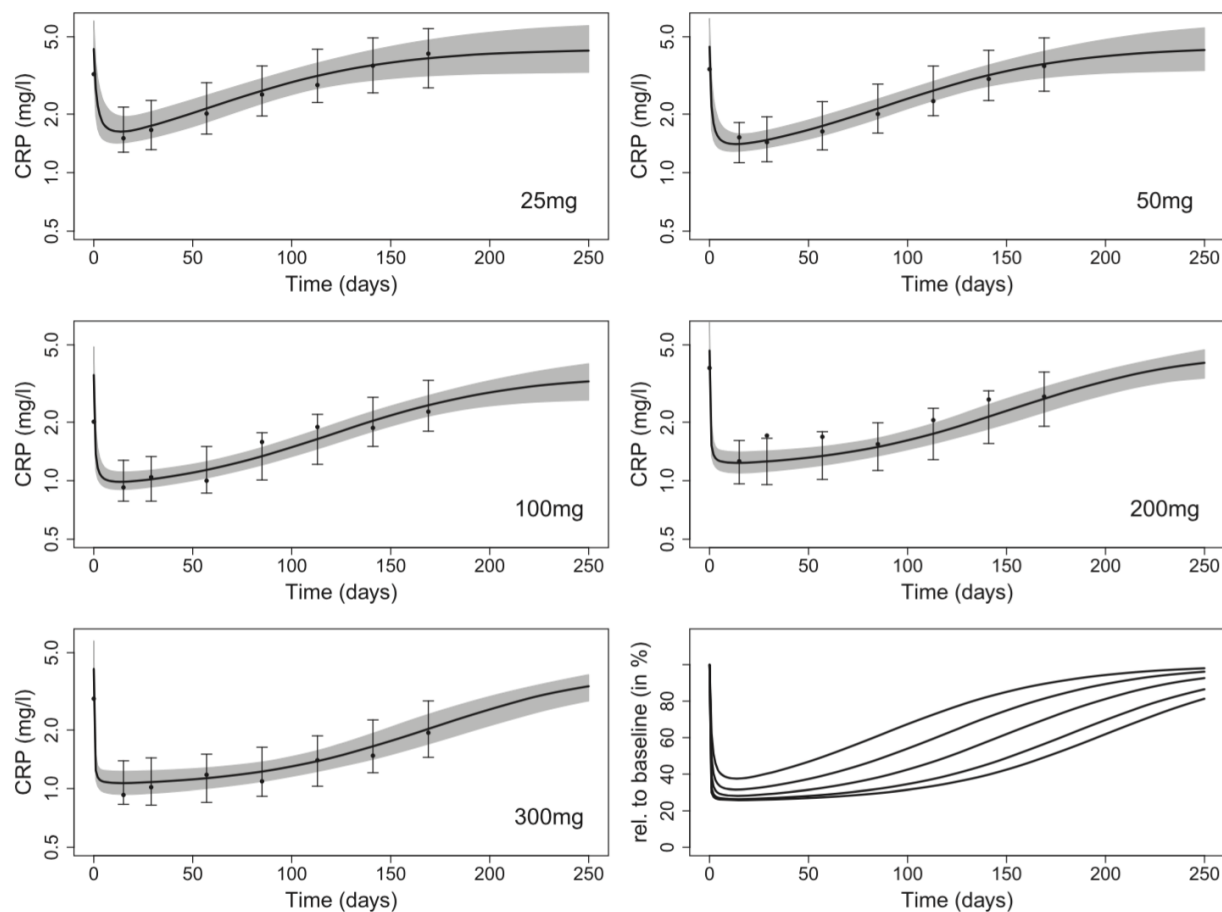
Figure 2. Bayesian analysis for the five single dose groups in the clinical trial. The dots represent the means of the CRP levels, the solid curves the median of the posterior CRP levels, the grey area the 95% posterior probability interval, and the vertical lines the 95% prediction probability interval. The last plot displays all posterior median curves relative to their baseline value.

Figure 3 is a comparison of the data from the multiple dose group of the trial and the prediction from the dose-time-response model derived from the Bayesian analysis of the single-dose data. All of the data points are within the 95% prediction intervals, which implies a good accuracy of the dose-time-response model predictions of the response over time for the multiple dose treatment.
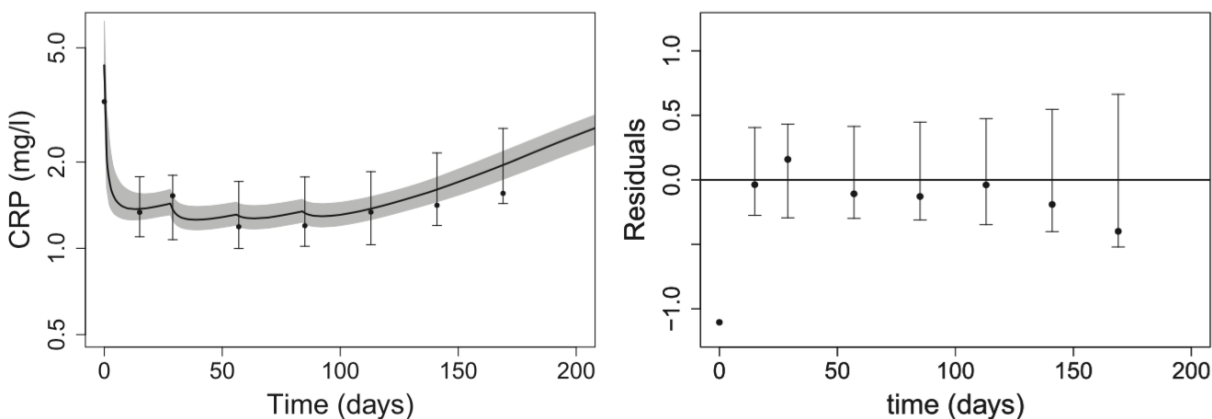
Figure 3. Bayesian prediction of the repeated dosing regimen based on the dose-time-response model derived from single-dose regimens

Figure 4 is a comparison of the fitted dose-time-response model with the standard fitted E_{max} (dose-response) model in estimating the difference to baseline. It is obtained by using Bayesian approach to fit the model for three time points at day 15, 85, and 169. Although the fits of data are similar for both models, the confidence band of the dose-time-response model is narrower than that of dose-response model, especially for the higher doses. This suggests that the predictions for higher doses in the dose-time-response model are more accurate on average.
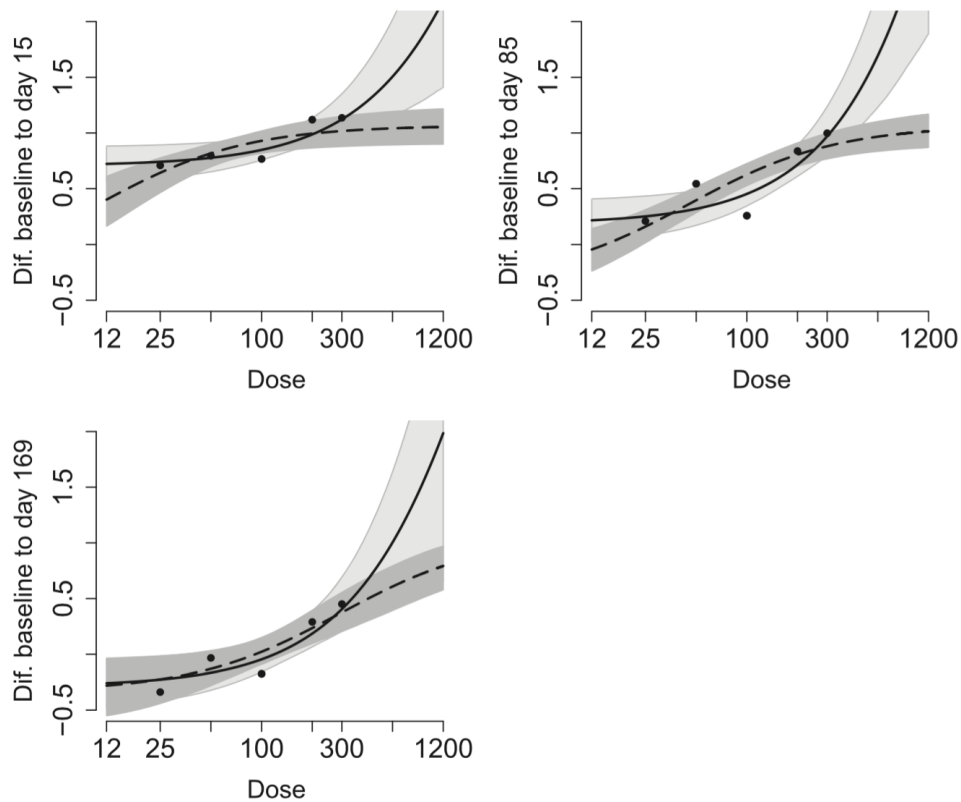
Figure 4 Comparison of the dose-time-response model with the standard E_{max} model. The dots represent the means of the measurements, the solid line the fitted E_{max} model, and the dashed line the fitted dose-time-response model. The light grey area and the dark grey area mark the respective 95%-confidence bands of the expected response.

6. Limitations

The followings are four major limitations of the paper:

- ML methods make inferences based on asymptotic theory, but it may not be valid for small samples in the experiments.
- In a Bayesian framework, unintended informative prior problems may occur because of inappropriate prior distributions for nonlinear models.
- Samples from the posterior are generated using MCMC methods, rather than direct calculations. Therefore, the estimation results are subject to the validity problem.
- In the clinical trail, only five single-dose arms and one multiple-dose arm are carried out for assessment, which may not be complete enough to achieve the accuracy of the model evaluation.

7. Issues and Challenges

The followings are three critical issues and challenges in terms of theories and technology:

- Form both technical and conceptual perspectives, It is often challenging to ensure the fitness of nonlinear regression models.
- Bayesian analysis requiring MCMC algorithms is computer-intensive and time-consuming. Doing so for many simulated datasets need a significant amount of computing powers and time.
- Convergence problems, identifiability problems, and ill-conditioning problems are commonly encountered in the experiments.

8. Replication

In the paper, MCMC algorithm is used for generated the posterior distribution of Bayesian Inference. However, the initial model is a nonlinear regression model having 5 parameters, which is challenging for us. So our simulation separate the data experience into two parts: using MCMC generate vectors and posterior distribution.

8.1 Markov Chain Monte Carlo (MCMC) Algorithm

The followings are R statements using MCMC algorithm to generate posterior distribution for a simple function:

```
library(mcmc)
h <- function(x) if (all(x >= 0) && sum(x) <= 1) return(1) else return(-Inf)
out <- metrop(h, rep(0, 5), 100)
out$accept
# increase accept rate
out <- metrop(out, scale = 0.01)
out$accept
# set the nbatch to 100
out <- metrop(out, nbatch = 1e2)
out$accept
plot(out$batch[ , 1])
# hist of posterior distribution
hist(out$batch[ , 1])
```
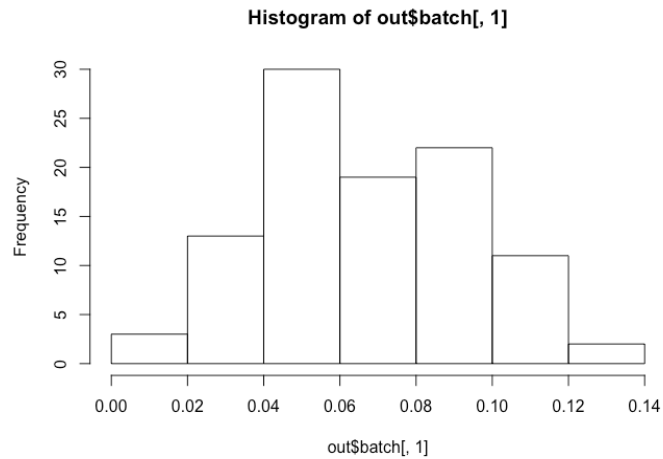
The plot of the 100 generated vectors is shown in Figure 5:
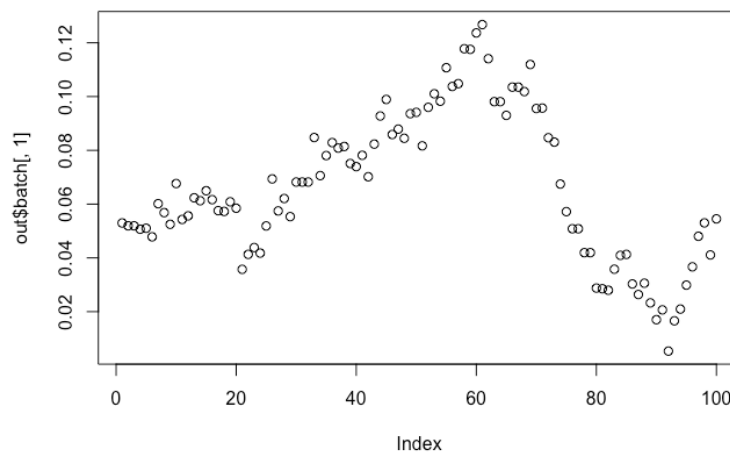
Figure 5 Plot of 100 vectors generated using Markov chain Monte Carlo (MCMC) method

The posterior distribution is shown in Figure 6:

Therefore, based on the posterior distribution, the estimation value is 0.07.

Histogram of out$batch[, 1]

Figure 6 Histogram of Posterior distribution

8.2 Maximum Likelihood (ML) method

Assume the estimated parameters from different samples are 123, 234, 345, 456, and 567.

To use Maximum Likelihood (ML) method to estimate the parameters of the model. The following R statements for are executed:

```
library(stats4)
f <- function(alpha){
  logL = n*log(alpha) - alpha*sum(x)
  return (logL)
}
x= c(123,234,345,456,567)
n = length(x)
a = optimize(f,c(0,1),maximum = TRUE)
1/a$maximum

## [1] 346.4544
```

The result shows that the estimated value of the parameter is 346.4544.

Reference

Hoos, A., Parmiani, G., Hege, K., Sznol, M., Loibner, H., Eggermont, A.,…Nichol, G. (2008). A Clinical Development Paradigm for Cancer Vaccines and Related Biologics. *Journal of Immunotherapy*. 30(1), 1-15.

Lange, M. L., & Schmidli, H. (2015). Analysis of clinical trials with biologics using dose–time-response models. *Statistics in Medicine,* 34, 3017–3028

Winthrop, K. L., Novosad, S. A., Baddley, J. W., L Calabrese, Chiller, T., Polgreen, P., … Smolen, J. (2015). Opportunistic infections and biologic therapies in immune-mediated inflammatory diseases: consensus recommendations for infection reporting during clinical trials and postmarketing surveillance. *Annals of the Rheumatic Diseases,* 74, 2107-2116