Incorporating Historical Data via Bayesian Analysis Based on The Logit Model

INCORPORATING HISTORICAL DATA VIA BAYESIAN ANALYSIS BASED ON THE LOGIT MODEL

BY

CHENXI YU, B.Eng.

A THESIS

SUBMITTED TO THE DEPARTMENT OF MATHEMATICAL & STATISTICS

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Master of Science (2018) McMaster University

(Mathematics & Statistics) Hamilton, Ontario, Canada

TITLE:	Incorporating Historical Data via Bayesian Analysis Based on The Logit Model
AUTHOR:	Chenxi Yu B.Eng., (Electrical Engineering) McMaster University, Hamilton, Canada
SUPERVISOR:	Dr. Narayanaswamy Balakrishnan
NUMBER OF PAGES:	ix, 46

This thesis is dedicated to my parents.

Abstract

This thesis presents a Bayesian approach to incorporate historical data. Usually, in statistical inference, a large data size is required to establish a strong evidence. However, in most bioassay experiments, dataset is of limited size. Here, we proposed a method that is able to incorporate control groups data from historical studies. The approach is framed in the context of testing whether an increased dosage of the chemical is associated with increased probability of the adverse event. To test whether such a relationship exists, the proposed approach compares two logit models via Bayes factor. In particular, we eliminate the effect of survival time by using poly-k test. We test the performance of the proposed approach by applying it to six simulated scenarios.

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Narayanaswamy Balakrishnan, for the patient guidance, encouragement and advice he provided throughout my time as his student. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so promptly. Without his guidance and persistent help, the completion of this thesis would have not been possible.

In addition, I would like to thank my family. For encouraging me in all of my pursuits and inspiring me to follow my dreams. I am especially grateful to my parents, who supported me emotionally and financially. I always knew that they believed in me and wishing the best for me. Thanks to my aunt, who teaches me to learn in life, to be happy, to understand and accept myself as I am!

To my friends, thank you for listening, offering me advice, and supporting me through the entire process. I can not forget the valuable conversation and suggestions of Jing Cai. I have often looked towards her for valuable suggestions, and she always helped me whenever I needed support in my research. I would like to thank Michael Li for his valuable help in R and JAGS.

Notation and abbreviations

EU	Experimental Unit
NTP	National Toxicology Program
BF	Bayes Factor
BFWO	Bayes Factor Without historical information
TIER	Type I Error Rate

Contents

Abstract **iv**

Acknowledgements **v**

Notation and abbreviations **vi**

1 Introduction **1**

2 Data description **3**

3 Background **5**

3.1 Poly-k test 5

3.2 Incorporating information from historical studies 8

4 Proposed method **10**

4.1 Bayesian inference 11

4.2 Bayes factor 13

4.3 Logit model 14

4.4 The logit function 17

4.5 Prior distributions 17

4.6 Bayes factor 19

5 Simulation and illustrative example **21**

5.1 Simulation study 21

5.2 Illustrative example 27

6 Concluding Remarks **29**

A R codes **31**

A Scenario 7 **43**

List of Figures

5.1 Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 1 wherein H_0 is true . 23

5.2 Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 2 wherein H_0 is true . 23

5.3 Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 3 wherein H_0 is true . 24

5.4 Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 4 wherein H_1 is true . 24

5.5 Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 5 wherein H_1 is true . 25

5.6 Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 6 wherein H_1 is true . 25

A.1 Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 7 wherein H_0 is true . 44

Chapter 1

Introduction

The National Toxicology Program (NTP) conducted a two-year bioassay experiment to determine whether the increased dosage of benzophenone is related to the increased occurrence of tumour. This bioassay experiment studied both sexes of two rodent species, and included one control group and three treatment groups. Each group consisted of 50 randomly chosen animals. For animals which died during the study period, they were screened for anomalies. At the end of the experiment, all surviving animals were sacrificed. The experimental unit (EU) ID, study period, survival time, dosage level and presence of tumour were recorded. However, the results of standard analyses of presence of tumour in female rats were not satisfactory. The records of female rats with tumour onset were 0 in the control group, 0 in the low-dose group, 1 in the mid-dose group and 2 in the high-dose group. The number with presence of tumour was extremely small. According to the report of NTP, the result of standard trend test shows that the p-value equals 0.074, which does not provide statistically significant effect at standard 5% level. Due to the high cost and long study period of bioassay experiments, repeating this experiment with increased number of EU in

each group is not realistic.

To solve this problem, Peddada et al.[2007] proposed a survival-adjusted test for detecting dose-related trends in tumor rates. This method incorporated data from historical control groups. These historical data arose from bioassay experiments conducted under conditions similar to this experiment. In this work, we propose a Bayesian analysis method comparing two logit models via Bayes factors. The method adopted is based on León Novelo et al.[2017] and Peddada's test. The goal of this method is to test whether an association between benzophenone and tumour rate exists or not.

Several authors have investigated the problem of incorporating historical data by Bayesian approach. Hoel and Yanagawa[1986] proposed a linear trend test for binomial reponse data by using beta prior distribution for the historical control information. Hobbs et al.[2012] proposed a Bayesian modification of the commensurate prior model. Unlike non-linear models in our approach, they incorporated patient-level historical data by using general and generalized linear mixed regression models.

This thesis is divided into six parts. In Chapter 2, we briefly described the datasets analyzed in this thesis. Chapter 3 outlines poly-k test and Peddada et al.'s test. We then introduce the Bayesian method and Bayes factors in Chapter 4. Chapter 5 presents results from simulation studies under different scenarios. Some concluding remarks are made in Chapter 6.

Chapter 2

Data description

Data that will be analyzed in this thesis are introduced in this chapter. Those data are arising from bioassay experiments carried out by National Toxicology Program. Usually, the duration for these experiments are two years, and they are designed to test whether certain chemical has any effect on the occurrence of tumour or not. All of the seven studies in our datasets have taken female rats as experimental unit. The data contain 7 groups altogether: six historical study groups and one current study group. Each record in the dataset consists of six variables: rat ID, study, dose level, survival time, study duration, and a dichotomous variable which indicates the existence of tumour or not. Rat IDs are different for rats in the same group, but it may be the same for two rats from different groups. The study variable indicates which study group the record comes from. The current study contains four dose levels: 0, 312, 625 and 1250 ppm(parts per million). The historical studies only have a zero-dose level. The survival time for each rat is recorded for each group. Due to the impact of some environmental factors, some rats may die before the end of the study. All died rats were screened for anomalies. As mentioned earlier, the bioassay experiments typically

run for 2 years. All study groups chosen in this dataset are from 2-year studies. Since the study duration is recorded in days, it may vary in different study groups.

The data for current study arose from a NTP technical report [2006] on study of benzophenone. This study is to determine the effects of benzophenone on rats and mice to identify potential toxics or carcinogenic hazards to humans. In the study, groups of 50 animals were fed diets containing 0, 312, 625 or 1,250 ppm benzophenone for 2 years. Tissues from more than 40 sites were examined for each animal. The data for six historical groups are from six previous bioassay experiments which were carried out under the same condition (caging, food, delivery method, etc.) as the benzophenone study considered now. The records for control groups are chosen. As the data shows, no rats developed tumour in control groups of historical studies. For the sake of consistency, only records from female rats are used in this thesis. The table below shows 7 rows of dataset as an example.

Table 1. Seven rows of the benzophenone dataset

Rat ID	Study	Dose	Survival time	Study duration	Tumor No=0, Yes=1
244	Current	0	288	731	0
288	Current	312	609	731	0
301	Current	625	528	731	1
385	Current	1250	485	731	1
126	H1	0	677	735	0
415	H2	0	728	729	0
144	H6	0	735	735	0

Chapter 3

Background

The problem that will be discussed in this thesis is whether the increased level of certain chemical treatment is associated with the increasing probability of presence of tumour. In a typical bioassay experiment, EUs are exposed to some chemical until they die or be sacrificed at the termination of the study period. In addition to the presence of tumour or not, survival time of EUs were also recorded. Here in our work, the cause of death is assumed to be unknown.

3.1 Poly-k test

Consider a bioassay experiment with I treatment groups and one control group. Animals in the ith group receive dose m_i, $i = 0, 1, \ldots, I-1$. Suppose there are J studies in the test, $j = 0, 1, \ldots, J-1$. The first group is the one from current study. The historical studies are indexed by $1, \ldots, J-1$. Let n_{ij} denote the number of individuals with the presence of tumor. In Cochran-Armitage linear trend test [1955], it is assumed that all EUs are at equal risk of developing the tumour during

the study period. Therefore, the tumour rate in the ith group is estimated as

$$r_{ij} = \frac{d_{ij}}{n_{ij}}, \tag{3.1}$$

where d_{ij} denotes the total number of animals that developed a tumour in the ith treatment group in study j and n_{ij} denotes the total number of animals assigned to the ith treatment group in study j. However, if animals die early in a bioassay experiment, they are obviously at lower risk of developing a tumour compared to those EUs who live to the end of experiment. In other words, those animals who die early and do not developed a tumour make less contribution to the survival count. One way of correcting this problem has been introduced by Bailer and Portier[1988]. In this method, the total number at risk for test is defined as

$$n^*_{ij} = \sum_{l=1}^{n_{ij}} \omega_{ijl}, \tag{3.2}$$

where ω_{ijl} is a weight assigned to the lth animal in the ith group. The proportion is modified as

$$r_{ij} = \frac{d_{ij}}{n^*_{ij}}. \tag{3.3}$$

In poly-k test, the weight is set to be 1 if the corresponding animal dies with tumour and the weight is equal to the kth power of t_{ijl} over $t_{max,j}$, where t_{ijl} is the survival time of the lth animal in dose group i in study j and $t_{max,j}$ is the maximum survival time in study j. In most cases, $t_{max,j}$ equals the study duration of study j. The weight

ω_{ijl} can then be expressed as

$$\omega_{ijl} = \begin{cases} 1 & \text{if the } l\text{th animal in the } i\text{th group in study } j \text{ dies with tumour} \\ (\frac{t_{ijl}}{t_{max,j}})^k & \text{otherwise.} \end{cases} \tag{3.4}$$

The weights ω_{ijl} are all set to 1 for the Cochran-Armitage test. Kupper et al.[1986] found that most presences of tumour seem to occur at the rate of a third to fifth order power in time. As recommended by Bailer and Portier[1988], the value of k is set to 3. The method in this thesis is built on such a poly-3 test. Let π_{ij} denote the tumour rate for animals surviving until the end of the jth study in the ith group. The poly-k test suggests that the probability that a EU develops the tumour in the duration of the study is proportional to the weight assigned to that EU. Let z_{ijl} represent the quantal response, where it is set to 1 if the lth EU in the ith group in study j develops the tumour, and it is equal to 0; if it is not the case, that is,

$$P(z_{ijl} = 1) = \omega_{ijl}\pi_{ij}. \tag{3.5}$$

The mean of the total number of EU who exhibit tumour is then

$$\begin{aligned} E(\sum_l z_{ijl}) &= \sum_l P(z_{ijl} = 1) \\ &= \sum_l \omega_{ijl}\pi_{ij} \\ &= \pi_{ij} \sum_l \omega_{ijl} \\ &= \pi_{ij} n^*_{ij}. \end{aligned} \tag{3.6}$$

Note that d_i is the observed value of $E(\sum_l z_{ijl})$, and $r_{ij} = \frac{d_i}{n^*_{ij}}$. In other words, r_{ij} is the moment estimate of π_{ij}. The survival time is incorporated into the analysis by modifying n_{ij} to n^*_{ij}.

3.2 Incorporating information from historical studies

In bioassay experiments, the occurrence of tumours is very rare. However, these bioassay experiments are time-consuming and costly. To increase the power of the analysis, a rich collection of information is needed. So, to overcome this problem, a method of incorporating historical data is needed. In the past, several methods have been proposed to incorporate historical data. However, most of these methods do not adjust for survival. Peddada's test [2007] can correct this problem. In Peddada's test, poly-3 adjustment has been applied in their analysis. Multiple historical control groups database are accounted in their analysis. All these historical studies have the same sort of exposure (feed, gavage, inhalation, skin painting, etc.) as the study considered here. A restriction has been placed on these historical study groups: no genetic drift. The test statistic proposed by Peddada et al.[2007] is $W = max(D_1, D_2)$, where D_1 represents the distance between the current control group and the current treatment groups, and D_2 represents the distance between historical control groups and current treatment groups. These statistics are motivated by Bieler-Williams test and Peddada-Kissling test, and as usual a significance level α is set. The $W_{1-\alpha}$ quantile is estimated to compare with test statistic W value. If the observed value of

W is greater than $W_{1-\alpha}$, the null hypothesis would be rejected.

Chapter 4

Proposed method

The method proposed in this thesis compares two models via Bayesian analysis by the use of Bayes factors. One model is built under the null hypothesis, while the other one supports the alternative hypothesis. The model M_1 assumes tumour rates are related to dose levels, while model M_0 assumes that different dose levels have no effect on tumour rates. The Bayes factor is then estimated to compare these two models. According to the Bayes factor, one model is chosen over the other. For example if model M_1 is chosen, it means the hypothesis corresponding to M_0 is rejected. In this chapter, we will introduce the method of estimating Bayes factors and briefly describe the two logit models considered.

4.1 Bayesian inference

Suppose there are two random variables, X and Y, having a joint PDF $f_{X,Y}(x, y)$. Then, the marginal density of X is

$$f_X(x) = \int f_{X,Y}(x, y)\ dy. \tag{4.1}$$

Then, the conditional density of Y, given $X = x$, is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}. \tag{4.2}$$

Similarly, the conditional distribution of X, given $Y = y$, is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \tag{4.3}$$

Then, we evidently have

$$f_{X,Y}(x, y) = f_{Y|X}(y|x)f_X(x) = f_{X|Y}(x|y)f_Y(y). \tag{4.4}$$

In classical Bayesian analysis, all unknown parameters are treated as random variables. Let θ denote the unknown parameters. Suppose θ has a probability distribution $f_\Theta(\theta)$. This assumption is based on our previous knowledge about parameter θ. The distribution $f_\Theta(\theta)$ is called the prior distribution. Let Z denote the data we observed, and $f_{Z|\Theta}(z|\theta)$ be the conditional distribution of Z, given $\Theta = \theta$. The joint distribution of Θ and Z is clearly

$$f_{Z,\Theta}(z, \theta) = f_{Z|\Theta}(z|\theta)f_\Theta(\theta). \tag{4.5}$$

Then, we have the marginal distribution of Z to be

$$f_Z(z) = \int f_{Z,\Theta}(z,\theta)d\theta = \int f_{Z|\Theta}(z|\theta)f_\theta(\theta)d\theta. \tag{4.6}$$

The conditional distribution of θ, given $Z = z$ is then

$$f_{\Theta|Z}(\theta|z) = \frac{f_{Z,\Theta}(z,\theta)}{f_Z(z)} = \frac{f_{Z|\Theta}(z|\theta)f_\Theta(\theta)}{\int f_{Z|\Theta}(z|\theta)f_\Theta(\theta)d\theta}. \tag{4.7}$$

This conditional distribution represents the posterior distribution of Θ, given $Z = z$.

The Bayesian analysis, attributed originaly by Jeffery [1935], is a statistical method which combines prior information about parameters with information from data to get a statistical inference. There are three factors that play important roles in Bayesian method [2014]. The first one is the prior probability distribution, which represents all previous knowledge about parameters before seeing the data. The second factor is the observed data, which represents observed evidence from real life. This evidence is expressed by the likelihood function of the data. The likelihood function gives the probability of the data given parameters. The third one combines information from previous two factors, which is called the posterior distribution. Posterior distribution combines prior distribution and observed data via Bayes' theorem.

The conditional distribution of Z, given $\Theta = \theta$, is also the likelihood function of the data. From (4.7), we can find

$$f_{\Theta|Z} \propto f_{Z|\Theta}(z|\theta)f_\theta(\theta). \tag{4.8}$$

Since the marginal distribution $f_Z(z)$ is constant with respect to Θ, it is easy to see that the posterior density is proportional to the product of the likelihood function and the prior density.

4.2 Bayes factor

The Bayes factor is a ratio of the probability of data under two models. These two models usually correspond to the null and alternative hypothesis [Good and Hardin, 2012]. Suppose we have a set of observed data D, model M_0 corresponding to the null hypothesis H_0, and model M_1 corresponding to the alternative hypothesis H_1. The likelihood probability of data under M_0 is given by

$$P(D|M_0) = \frac{P(M_0|D)P(D)}{P(M_0)}. \tag{4.9}$$

Similarly, the likelihood probability of data under M_1 is given by

$$P(D|M_1) = \frac{P(M_1|D)P(D)}{P(M_1)}. \tag{4.10}$$

Then, the Bayes factor B_{01} is estimated as

$$\begin{aligned} B_{01} =& \frac{P(D|M_0)}{P(D|M_1)} = \frac{P(M_0|D)P(D)}{P(M_1|D)P(D)} \frac{P(M_1)}{P(M_0)} \\ =& \frac{P(M_0|D)P(M_1)}{P(M_1|D)P(M_0)} \end{aligned} \tag{4.11}$$

The most common interpretation for Bayes factor has been given by Jeffery [1935]. It is described in the table below:

Table 2: Jefferys' scale interpretation of the Bayes Factor	
Bayes factor B_{01}	Interpretation
10-30	Strong evidence for H_0
3-10	Moderate evidence for H_0
1-3	Anecdotal evidence for H_0
1	No evidence
1/3-1	Anecdotal evidence for H_1
1/3-1/10	Moderate evidence for H_0
1/10-1/30	Strong evidence for H_0

Note, that $B_{10} = 1/B_{01}$.

4.3 Logit model

Suppose there are I treatment groups in the current study, $i = 0, \ldots, I-1$, and the associated dose levels are $d_0 < d_1 < \cdots < d_I$, with d_0 standing for dose level in the control group and $d_0 = 0$. Recall the dataset described in Chapter 2, where besides the control group, there are 3 treatment groups in the study: low-dose, middle-dose and high-dose groups. Thus, for the current study, $I = 4$. However, in historical studies, there are only control groups. Let J be the total number of studies in the dataset, $j = 0, \ldots, J-1$. In the dataset described in Chapter 2, there is one current study and six historical studies. Let $j = 0$ index the current study. Let π_{ij} denote the tumour rate in the ith group in the jth study. The null hypothesis assumes that

the tumour rates are the same among different dose groups; that is

$$H_0 : \pi_{00} = \pi_{10} = \cdots = \pi_{(I-1)0}. \tag{4.12}$$

Since historical studies do not contain any information on treatment groups, we only discuss the case for current study here. The alternative hypothesis states that there is a positive trend for tumour rates in treatment groups, that is,

$$H_1 : \pi_{00} \leq \pi_{10} \leq \cdots \leq \pi_{(I-1)0}, \tag{4.13}$$

where at least one inequality in H_1 is strict.

The model M_1 is built on the hypothesis H_1. The model M_1 is specified as

$$M_1 : z_{ijl} \sim Bernoulli(\omega_{ijl}\pi_{ij}), \tag{4.14}$$

where

$$logit(\pi_{ij}) = \log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \alpha + X_i^{'}\gamma + Y_j^{'}\beta, \tag{4.15}$$

z_{ijl} stands for the binary response which indicates the existence of tumour in the lth animal in the ith treatment group in the jth study, and ω_{ijl} is the weight assigned by the poly-3 test. The details about ω_{ijl} was discussed earlier in Chapter 3. Since z_{ijl} is a binary quantity, it is common to assume that it has a Bernoulli distribution as in (4.14). In the link function in (4.15), α denotes the environmental effect on tumour rate, γ denotes the dose effect, and β denotes the difference across studies, and γ and β are vectors, given by

$$\gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \dots & \gamma_{I-1} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_{j-1} \end{bmatrix}, \tag{4.16}$$

X_i is a $I-1$ dimensional vector with its first i entries equal to 1, and all the following entries are equal to 0. For the case in Chapter 2, with $I = 4$, we have

$$X_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \ X_2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}, \ X_3 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \tag{4.17}$$

Thus, we have $X_1^{'}\gamma = \gamma_1$ representing the effect of low dose, $X_2^{'}\gamma = \gamma_1+\gamma_2$ representing the effect of medium dose, and $X_3^{'}\gamma = \gamma_1+\gamma_2+\gamma_3$ representing the effect of high dose. This model incorporates the same assumptions as in Peddada's method. Peddada's model assumes there is a positive trend in the effects of dose groups on tumour rates. under this assumption, we have $\gamma_1 \leq \gamma_1+\gamma_2 \leq \gamma_1+\gamma_2+\gamma_3$. In other words, all entries in the vector γ must be non-negative. The vector U_j indicates the corresponding study. The product of U_j and β represents the small deviations of the tumour rate for historical control groups from the current control group. In vector U_j, all entries are set to 0, except the $(j-1)$th entry. For example, $U_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$ for $J = 6$.

In contrast, to the alternative model M_1, the null model assumes that the chemical dose does not have an impact on tumour rates; that is,

$$M_0: \ z_{ijl} \sim Bernoulli(\omega_{ijl}\pi_{ij}) \ with \ logit(\pi_{ij}) = \alpha + U_j^{'}\beta. \tag{4.18}$$

4.4 The logit function

In a linear probability model, the probability of an event will be a linear function of covariates, of the form

$$\pi_{ij} = x_i^{'}\beta, \tag{4.19}$$

where β is a vector of regression coefficients and x_i is a vector of observed covsriates. The left hand side of this equation is a probability, and so its value must be in the range (0,1). The right hand side of this equation can take any value on the real line, and so, the prediction value may be out of the range (0,1). For this reason, we consider logit model of the form

$$\text{logit}(\pi_{ij}) = \log(\frac{\pi_{ij}}{1-\pi_{ij}}) = x_i^{'}\beta. \tag{4.20}$$

The logit function maps π_{ij} from the range (0,1) to $(-\infty, \infty)$, and so we get

$$\pi_{ij} = \frac{\exp(x_i^{'}\beta)}{1+\exp(x_i^{'}\beta)}, \tag{4.21}$$

which can guarantee π_{ij} to stay in the range (0,1).

4.5 Prior distributions

The parameter α stands for the environment effect on tumour rates. We assume it is the same for all groups in all studies. α is set to have a normal prior distribution with mean $\mu_\alpha = 0$ and variance $\sigma_\alpha = 1$. The vector β stands for difference due to

study groups. Since the environment and units in experiment are different for different studies, these factors may cause a variation in the tumour rate. The vector β is assumed to have a multivariate normal distribution with mean μ_β equaling 0 and variance-covariance matrix $\sigma_\beta^2 I_{J-1}$. We assume no relations exists between different studies. It is reasonable to set all covariates to be 0. The parameter σ_β measures the difference of effect due to studies. If the value of σ_β is close to 0, we can say historical studies do not effect the prediction of tumour rate. In contrast, if σ_β has a nonzero value, it indicates there is a difference between tumour rates in current study and historical study groups. We assume that the effect due to environment factor and study groups are not significant, so that their effect can be ignored. So, we set the mean of α and elements in β to be 0.

In the alternative hypothesis, we assume the higher the dose level, the greater the chance of the animal developing a tumour. The vector γ is assumed to have a truncated normal distribution. Due to their positive effect on tumour rate, all entries in γ has to be non-negative. Thus, we have the following prior set-up:

$$\begin{aligned} \alpha &\sim N_1(\alpha|\mu_\alpha = 0, \sigma_\alpha^2 = 1), \\ \beta &\sim N_{J-1}(\beta|\mu_\beta = 0, \sigma_\beta^2 I_{J-1}), \\ \sigma_\beta &\sim TN_{j-1}^+(0, \lambda_\beta^2 = 1), \\ \gamma &\sim TN_{I-1}^+(\mu_\gamma = 0, \sigma_\gamma^2 I_{I-1}), \\ \sigma_\gamma &\sim TN^+(0, \lambda_\gamma^2 = I). \end{aligned} \tag{4.22}$$

Here, N_m denotes a m-dimensional multivariate normal distribution, and when $m = 1$, N_1 simply denotes a uni-variate normal distribution; TN^+ denotes the truncated normal distribution with all entries have been restricted to be positive.

4.6 Bayes factor

In this approach, the Bayes factor will be estimated by using MCMC methods, as follows:

$$\begin{aligned} BF_{01} =& \frac{\int P_0(z|\theta_1)d\theta_1}{\int P_1(z|\theta_1,\theta_0)P(\theta_0)P(\theta_1)d\theta_0 d\theta_1} \\ &= \frac{E(P_0(z|\theta_1))}{E(P_1(z|\theta_1,\theta_0))} \\ &= \frac{1}{T}\sum_{t=1}^{T}\frac{P_0(z|\theta_1)}{P_1(z|\theta_0,\theta_1)}. \end{aligned} \tag{4.23}$$

Here, $\theta_1 = (\gamma, \sigma_{gamma})$ and $\theta_0 = (\alpha, \beta)$,and

$$P_0(z|\theta_1) = \prod_{ijl} Bernoulli(z_{ijl}|\omega_{ijl}\pi^0_{ij}) \tag{4.24}$$

with π^0_{ij} =logit$(\alpha + U_j'\beta)$,

$$P_1(z|\theta_1,\theta_0) = \prod_{ijl} Bernoulli(z_{ijl}|\omega_{ijl}\pi^1_{ij}) \tag{4.25}$$

with π^1_{ij} =logit$(\alpha + X_i'\gamma + U_j'\beta)$.

The values of θ_1 and θ_0 will be estimated by Gibbs sampler. Gibbs sampler is a method which simulates from the posterior distribution of θ [Albert and Chib, 1993]. Suppose there are k components in the parameter vector θ,

$$\theta = (\theta_1, \theta_2, \ldots, \theta_k). \tag{4.26}$$

To implement Gibbs sampler, the initial values for θ are necessary. Usually, the initial value is set based on previous experience. Let θ_i denote values after ith iteration and

θ_0 denote initial values. Then, θ_i values are simulated according to their conditional distributions:

$$\begin{aligned} \theta_1^{(1)} \text{ from } P(\ \theta_1 |\ \theta_j^{(0)}, j \neq 1), \\ \theta_2^{(1)} \text{ from } P(\ \theta_2 |\ \theta_1^{(1)},\ \theta_j^{(0)}, j > 2), \\ \vdots \\ \theta_k^{(1)} \text{ from } P(\ \theta_k |\ \theta_j^{(1)}, j < k). \end{aligned} \tag{4.27}$$

After t iterations, the sample $\theta_t = (\theta_1^{(t)}, \ldots, \theta_k^{(t)})$ is obtained. As t approaches to infinity, the joint distribution of θ_t approaches the joint distribution of θ. In the simulation study of this thesis, JAGS was used to implement this sampling process.

Chapter 5

Simulation and illustrative example

5.1 Simulation study

In the simulation study, six sets of data were created by pseudo-random sampling from known probability distributions. In the simulated datasets, all covariate values were taken to be the same as in the dataset described in Chapter 2. These datasets have one control group and three treatment groups in current study, and one control group for each historical study. Each dataset corresponds to one scenario. There are six scenarios in the simulation study. As in the benzophenone dataset described in Chapter 2, the tumour rates are set to be low in all scenarios.

For the cases discussed in the first three scenarios, H_0 is the true hypothesis. In Scenario 1, tumour rates are set to be 0.05 for all dose groups and studies. In Scenario 2, tumour rates are set to be 0.05 plus a study effect. This study effect has a uniform distribution with lower endpoint -0.03 and upper endpoint 0.03. Thus, in this scenario, the study effect has a mean of 0. Scenario 3 is similar to Scenario 2, except that the study has a uniform distribution with lower endpoint 0 and upper endpoint 0.05. In

Scenario 3, the study effect is always non-negative.

In contrast, the last three scenarios support hypothesis H_1. For treatment groups in the current study, the tumour rate is increasing with the dose level. Thus, we have the following:

Scenario 1: $\pi_{ij} = 0.05$ for all treatment groups and studies,

Scenario 2: $\pi_{ij} = 0.05 + u_j$ with $u_j \sim$ Uniform(-0.03,0.03), for $j = 2,\ldots,J$,

Scenario 3: $\pi_{ij} = 0.05 + u_j$ with $u_j \sim$ Uniform(0, 0.05), for $j = 2, \ldots, J$,

Scenario 4: $\pi_{ij} = 0.05 + i(0.04)$,

Scenario 5: $\pi_{ij} = 0.05 + i(0.04) + u_j$, with $u_j \sim$ Uniform(-0.03,0.03), for $j = 2, \ldots, J$,

Scenario 6: $\pi_{ij} = 0.05 + i(0.04) + u_j$, with $u_j \sim$ Uniform(0, 0.05), for $j = 2, \ldots, J$.

1000 datasets were generated from each scenario. The Gibbs-sampling method was used to compute the Bayes factors. Here, both the Bayes factor with the historical study groups data and Bayes factor without the historical study group data were calculated. $T = 10^3$ set of imputed parameters were generated for each model. The imputed parameter values were saved every ten iterations after a burn-in period of 10^3 iteration.

According to Jefferys' table presented in Section 4.2, we select model M_1 over M_0 when the Bayes factor BF_{10} is greater than 3, and we select model M_0 over M_1 when the Bayes factor BF_{10} is less than 3. The criterion is 3. Figures 1-6 show the plots of the Bayes factors for each scenario. BF_{10} stands for the Bayes factors which incorporate historical studies' data. $BFWO_{10}$ stands for the Bayes factors which do not include historical studies' data in the model. For visual convenience, we have used the log scale and the solid line on the plot corresponds to the value log(3).

Figure 5.1: Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 1 wherein H_0 is true

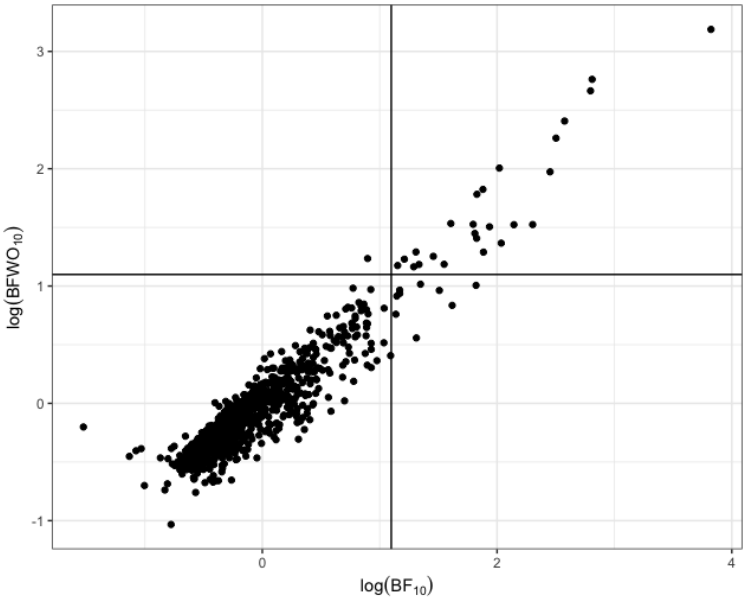


Figure 5.2: Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 2 wherein H_0 is true

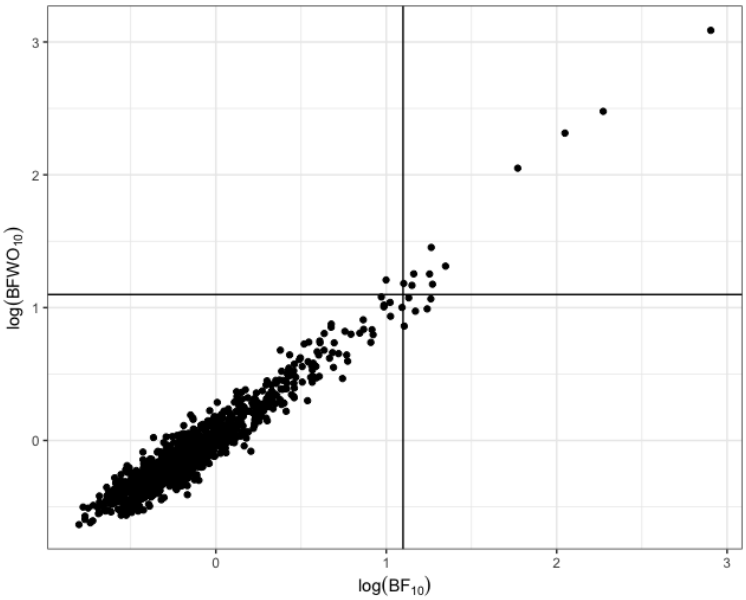


Figure 5.3: Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 3 wherein H_0 is true

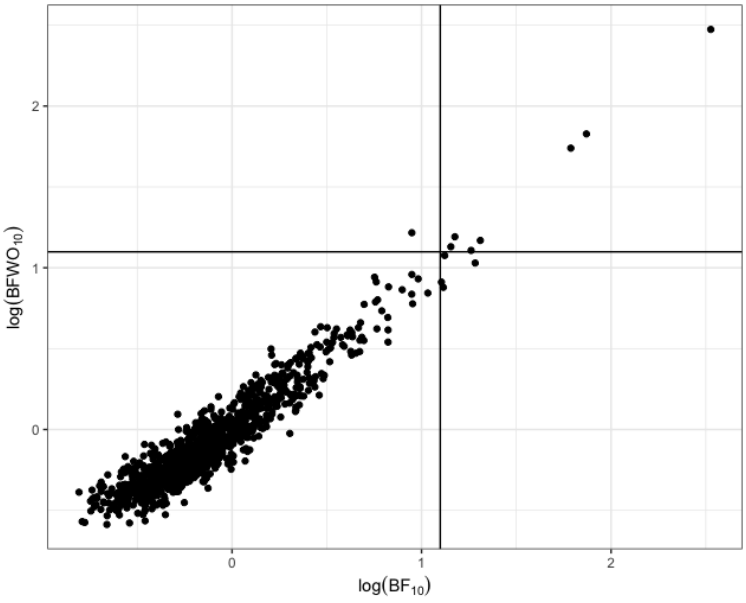


Figure 5.4: Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 4 wherein H_1 is true

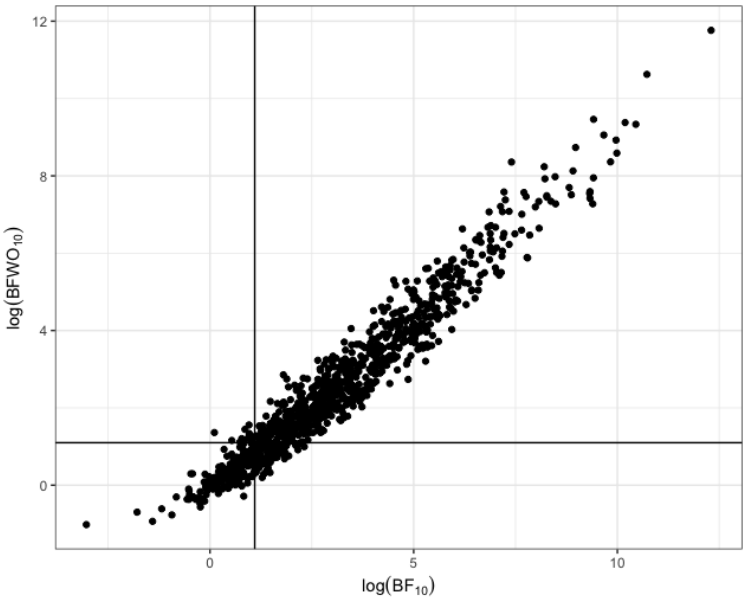


Figure 5.5: Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 5 wherein H_1 is true

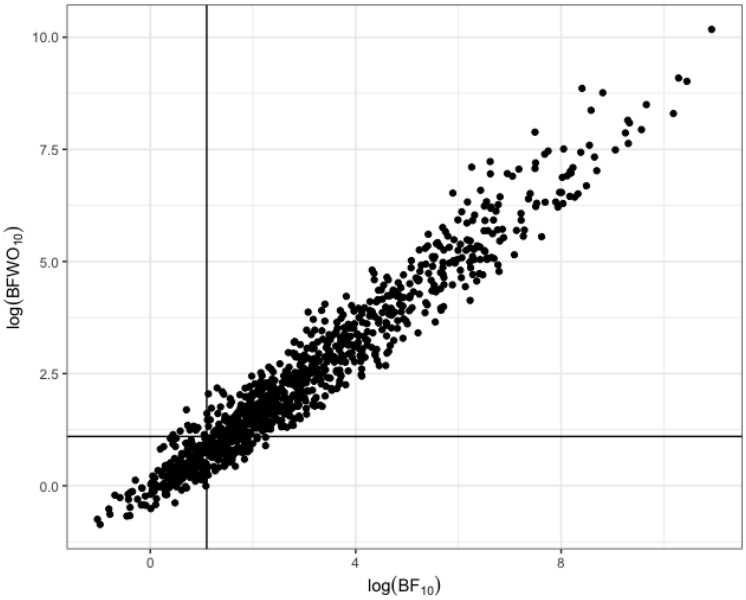
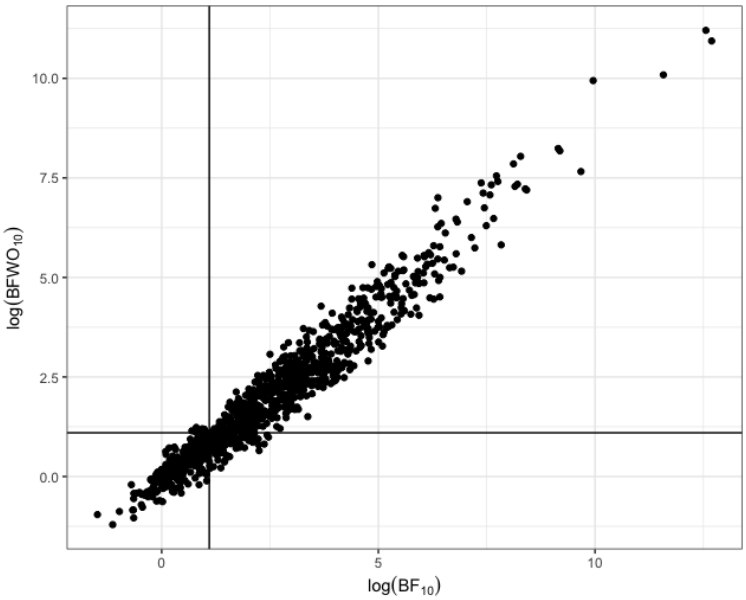


Figure 5.6: Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 6 wherein H_1 is true



The table below shows the performance of the proposed method. We calculated Type I error rates (TIER) for the first three scenarios and powers for the last three scenarios. Here again, BF stands for Bayes factors which incorporate historical control data in the analysis, while BFWO stands for Byes factors which do not incorporate historical control data in the analysis.

Table 3: TIER & Power for Proposed Bayesian approach

Scenario #	TIER for BF	TIER for BFWO	Power for BF	Power for BFWO
1	0.004	0.026	-	-
2	0.016	0.012	-	-
3	0.011	0.008	-	-
4	-	-	0.820	0.735
5	-	-	0.807	0.724
6	-	-	0.750	0.653

As we can be seen from Table 3, in the first three scenarios, the proposed approach controls type I error at the nominal level (which is 0.05). For scenario 1, the results show less TIER upon incorporating historical control data. For scenarios 2 and 3, TIERs for BF is a little bit higher than TIER for BFWO. Thus, TIERS for Scenarios 2 and 3 are higher than TIER for Scenario 1. In Scenarios 2 and 3, historical studies are assumed to have an effect on tumour rates. Powers for the last three scenarios are calculated. We can note that in all these three scenarios, the method incorporating historical control data is more powerful than the one which does not. The power in Scenario 4 is higher than in Scenario 5, which in turn is higher than in Scenario 6. When hypothesis H_1 is correct, study effect affects the performance of the proposed

method. When average tumour rates are more similar across different study groups, the method demonstrates higher power.

5.2 Illustrative example

With the simulated data, we also examine the data from the benzophenone study which has been discussed earlier in Chapter 2. We applied the same method as in the simulation study. We built two logit models and generated $T = 10,000$ posterior samples for each model. We saved the parameter values every 10 iterations after a 1,000 iteration burn-in. We obtained the value of $BF_{10} = 13.01$. This provides a strong evidence for hypothesis H_1. Therefore, we choose model M_1 to estimate the tumour rate for benzophenone study and the table below shows the estimated tumour rates.

Table 4: Estimated tumour rates for benzophenone study

Study	Tumour counts	Estimated tumour rate	2.5% quantile	97.5% quantile
C0	0/50	0.016	0.006	0.041
C1	0/50	0.020	0.008	0.047
C2	1/50	0.026	0.011	0.060
C3	2/50	0.035	0.013	0.095
H1	0/50	0.004	0	0.036
H2	0/50	0.006	0	0.040
H3	0/50	0.006	0	0.036
H4	0/50	0.006	0	0.027
H5	0/50	0.006	0	0.035
H6	0/50	0.008	0	0.034

In the benzophenone study data, there is one rat in the middle dose group and two rats in the high dose group who developed tumour. The tumour rate is approximate by 0.02 for middle dose group and 0.04 for the high dose group, and 0 for all other groups. Our estimated results show 0.026 for middle dose group and 0.035 for the high dose group, which is close to the observed estimates. However, estimation for the control group and the low dose group are relatively high. The estimated tumour rate is 0.016 for the control group and 0.020 for the low dose group. The estimation for historical control groups are small, which are in conformance with the fact that there is no rats developed tumour in these groups.

Chapter 6

Concluding Remarks

In this thesis, we have proposed a Bayesian approach to analyze data from bioassay experiments via Bayes factors. In this method, we compare two logit models corresponding to two hypothesis. In Chapter 2, we have described data which has been used to examine our proposed method. Chapter 3 briefly introduces poly-k test and Peddada's test which are used in our method. In Chapter 4, we have described our method in detail. We have demonstrated the proposed approach via a simulation study. We have examined the performance of our approach through six pseudo scenarios and also applied the method to a real data from benzophenone study. As the result demonstrates in simulation study, our method exhibits low type I error rate compared to the nominal level of 5%, and also demonstrates high power to reject the null hypothesis. In the application of the proposed method on data from Benzophenone study, we find our method to yield good estimates for tumour rates in different study groups.

In our analysis, we have taken $k = 3$ in the estimation of weights in the poly-k test. To extend this model, we can treat k as an unknown parameter, and determine a

suitable k for a given data. Due to the time constraint, we ran two chains for each scenario in our simulation, and so the result may not reveal the effect of initial values. It is possible to get more accurate result by running multiple chains. Furthermore, in our logit model, we only take into account the effects of historical studies, and ignored the impact of other factors. Future work could focus on the inclusion of all factors.

Appendix 1 A

R codes

```
# all packages required
library (R2jags)
library (readr)
library (tidyverse)
library (readxl)
library (dplyr)
library (MCPAN)

# initiate two empty vectors to store output
bfa <- c()
bfc <- c()

for (iter in 1:1000){
niter <- 11000 # number of iteration for Gibbs sampling
```

```
nburnin <- 1000 # number of burn in
nthin <- 10 # save every 10 values
# simulating data
ID <- seq (1, 500, by=1) # simulate rat ID
# simulate study group
Study <- c( rep("Current", 200), rep("H1", 50), rep("H2", 50), rep("H3",
x1 <- as.factor(Study)
x2 <- as.numeric(x1)
# simulate dose level
c_dose <- c( rep(0, 50), rep(312, 50), rep(625, 50), rep(1250, 50))
Dose <- c(c_dose, rep(0, 300))
y1 <- as.factor(Dose)
y2 <- as.numeric(y1)
y3 <- y2-1 #effect of Dose
# simulate rats' survival time
s <- runif(500, min=305, max=737)
Survival_time <- ceiling(s)
# set study duration to 737 days for all studies
Study_duration <- c(rep(737, 500)
# set tumour rate
tumor_rate <- 0.05 + y3*(0.04)
# the binary response observation is generated according to tumour rate
Tumor <- rep(1, 500)
for (m in 1:500){
```

```
  Tumor[m] <- rbinom (1, 1, tumor_rate[m])
}
# save data as a dataframe
dat <- data.frame(ID, Study, Dose, Survival_time, Study_duration, Tumor)

#using Poly-3 method to compute weight omega
     weight <- rep(1, length(Tumor))
     n <- length(Tumor)
     Y <- sum(Tumor)
     for (i in which(!Tumor)) {
       weight[i] <- (Survival_time[i]/Study_duration[i])^3
     }
my_omega <- weight
# omega_current is the weight for current study groups
omega_current <- weight[1:200] # omega values for current study groups
# omega_other is the weight for historical studies
omega_other <- weight[201:500] # omega values for historical study group

jags_data <- list(current = dat\$Tumor[1:200]
                  , other = dat\$Tumor[201:500]
                  , omega_current = omega_current
                  , omega_other = omega_other
                  , zero = 0
                  , one = 1
```

```
)

# get posterior samples for H1, in which gamma !=0
getSamplesModelH1 <- function(data) {
  model <- "
  model{
  # priors
  alpha ~ dnorm(zero,one)

  ## only use gamma for current study group

  sigma_gamma ~ dnorm(zero,1)T(0,1) # set sigma_gamma to have a truncate
  sigma_gamma_sq <- exp(2*log(sigma_gamma))

  for (b in 1:3){
  gamma[b] ~ dnorm(zero, 1/sigma_gamma_sq) T(zero, )
  }

  gamma_prime[1] <- zero
  gamma_prime[2] <- gamma[1]
  gamma_prime[3] <- gamma[1] + gamma[2]
  gamma_prime[4] <- gamma[1] + gamma[2] + gamma[3]

  for(dose in 1:4){
```

```
d_current[dose] <- exp(alpha + gamma_prime[dose])
}
for(dose in 1:4){
y_current[dose] <- d_current[dose]/(1 + d_current[dose]) #compute y by
for(obs in 1:50){
prob_current[(dose-1)*50+obs] <- omega_current[(dose-1)*50+obs]*y_curr
current[(dose-1)*50+obs] ~ dbern(prob_current[(dose-1)*50+obs])
}
}

## only use beta=study group, for study group, they do not have does e

sigma_beta ~ dnorm(zero,one)T(zero,)
sigma_beta_sq <- exp(2*log(sigma_beta))

for (study in 1:6){
beta[study] ~ dnorm(zero, 1/sigma_beta_sq)
d_other[study] <- exp(alpha + beta[study])
}

for(study in 1:6){
y_other[study] <- d_other[study]/(1 + d_other[study])
for(obs in 1:50){
prob_other[(study-1)*50+obs] <- omega_other[(study-1)*50+obs]*y_other[
```

```
other[(study-1)*50+obs] ~ dbern(prob_other[(study-1)*50+obs])
}
}
} "

jags_init <- list(list(alpha = 0
                       , sigma_gamma = 0.001
                       , gamma = c(0.001,0.001,0.001)
                       , sigma_beta = 0.001
                       , beta = c(0.001,0.001,0.001,0.001,0.001,0.001)
                  , list(alpha = 0
                         , sigma_gamma = 0.001
                         , gamma = c(0,0.2,0.3)
                         , sigma_beta = 0.001
                         , beta = c(0.001,0.1,0.2,0.3,0.4,0.5))
)

s <- jags(data
          , inits = jags_init
          , parameters = c("prob_current"
                           , "prob_other"
                           , "alpha"
                           , "gamma", "sigma_gamma"
                           , "beta" , "sigma_beta"
```

```
            )
            , model.file = textConnection(model)
            , n.chains = 2
            , n.iter = niter
            , n.burnin= nburnin
            , n.thin = nthin
  )
  return(s)
  #summary(s)

}




# get posterior samples for H0, in which gamma = 0
getSamplesModelH0 <- function(data) {
  model <- "
  model{
  # priors
  alpha ~ dnorm(zero,one)

  for(dose in 1:4){
  d_current[dose] <- exp(alpha)
```

```
y_current[dose] <- d_current[dose]/(1 + d_current[dose]) #compute y by
for(obs in 1:50){
prob_current[(dose-1)*50+obs] <- omega_current[(dose-1)*50+obs]*y_curr
current[(dose-1)*50+obs] ~ dbern(prob_current[(dose-1)*50+obs])
}
}

## only use beta for historical study group, for study group, they do

sigma_beta ~ dnorm(zero,one)T(zero,)
sigma_beta_sq <- exp(2*log(sigma_beta))

for (study in 1:6){
beta[study] ~ dnorm(zero, 1/sigma_beta_sq)
d_other[study] <- exp(alpha + beta[study])
}

for(study in 1:6){
y_other[study] <- d_other[study]/(1 + d_other[study])
for(obs in 1:50){
prob_other[(study-1)*50+obs] <- omega_other[(study-1)*50+obs]*y_other[
other[(study-1)*50+obs] ~ dbern(prob_other[(study-1)*50+obs])
}
}
```

```
} "

jags_init <- list(list(alpha = 0
                       , sigma_beta = 0.001
                       , beta = c(0.001,0.001,0.001,0.001,0.001,0.001)
                  , list(alpha = 0
                         , sigma_beta = 0.001
                         , beta = c(0.001,0.1,0.2,0.3,0.4,0.5))
)

s <- jags(data
          #, inits = jags_init
          , parameters = c("prob_current"
                           , "prob_other"
                           , "alpha"
                           , "beta" , "sigma_beta"
          )
          , model.file = textConnection(model)
          , n.chains = 2
          , n.iter = niter
          , n.burnin=nburnin
          , n.thin = nthin
)
return(s)
```

```
  #summary(s)

}

a <- getSamplesModelH1(jags_data)

b <- getSamplesModelH0(jags_data)


# calculating BF by using median of samples
probc_H1 <- a$BUGSoutput$sims.list$prob_current
probo_H1 <- a$BUGSoutput$sims.list$prob_other
probc_H0 <- b$BUGSoutput$sims.list$prob_current
probo_H0 <- b$BUGSoutput$sims.list$prob_other
for (i in 1:200) {
  probc_H1[,i] <- (1-probc_H1[,i])*(1-Tumor[i])+probc_H1[,i]*Tumor[i]
}

for (i in 1:200) {
  probc_H0[,i] <- (1-probc_H0[,i])*(1-Tumor[i])+probc_H0[,i]*Tumor[i]
}

for (i in 1:300) {
  probo_H1[,i] <- (1-probo_H1[,i])*(1-Tumor[i+200])+probo_H1[,i]*Tumor[i
```

```
}
for (i in 1:300) {
   probo_H0[,i] <- (1-probo_H0[,i])*(1-Tumor[i+200])+probo_H0[,i]*Tumor[i
}

 rowProds(probc_H1)
 rowProds(probc_H1)
BFWO <- rowProds(probc_H1)/rowProds(probc_H0)
bfn <- rowProds(probc_H1)*rowProds(probo_H1)
bfd <- rowProds(probc_H0)*rowProds(probo_H0)
BF <- bfn/bfd
summary((BFWO))
bfa[iter] <- median(BF) #BayesFactor with all information

bfc[iter] <- median(BFWO)# BayesFactor without historical groups informa

}
medi4 <- data.frame(bfa,bfc)
 # print(lst)
 # sink()

## plot

library(ggplot2)
```

```
gg <- (ggplot(medi, aes(x=log(bfa), y=log(bfc)))
        + theme_bw()
        + labs(x=expression(log(BF[10])))
        + labs(y=expression(log(BFWO[10])))
        + geom_hline(yintercept = log(3))
        + geom_vline(xintercept = log(3))
        + geom_jitter()
      )

write.table(medi4,"/Users/yuc6/Desktop/simdata/medi.txt")
```

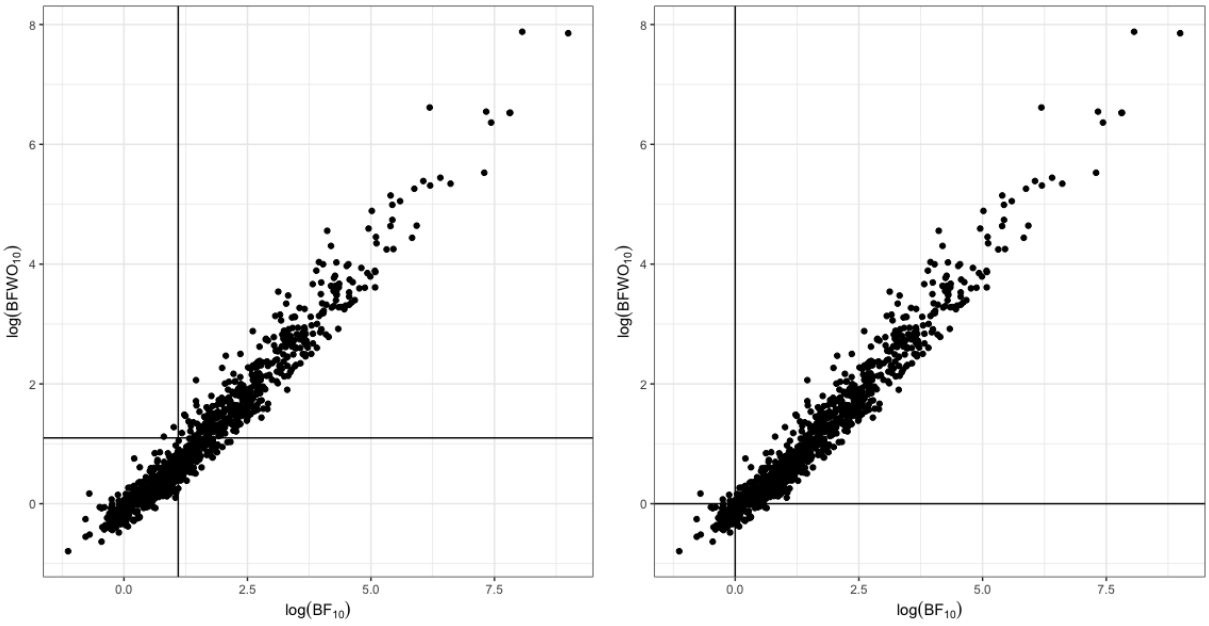
Appendix 2 A

Scenario 7

Besides the six scenarios discussed in Chapter 5, we simulate one more scenario to mimic the benzophenone example. That is,

Scenario 7: $\pi_{ij} = 0.02 + i(0.02)$.

Figure A.1: Bayes Factors BF_{10} and $BFWO_{10}$ for Scenario 7 wherein H_0 is true



The result are shown in Figure A.1. Under scenario 7, the power is 0.574 for BF and 0.439 for BFWO. If we change the criterion value from log(3) to log(1), the power becomes 0.919 for BF and 0.871 for BFWO.

Bibliography

J. H. Albert and S. Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

P. Armitage. Tests for linear trends in proportions and frequencies. *Biometrics*, 11 (3):375–386, 1955.

A. J. Bailer and C. J. Portier. Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics*, 44(2): 417–431, 1988.

P. I. Good and J. W. Hardin. *Common Errors in Statistics (and How to Avoid Them)*. John Wiley & Sons, Hoboken, New Jersey., 2012.

B. P. Hobbs, D. J. Sargent, and B. P. Carlin. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis (Online)*, 7(3):639, 2012.

D. Hoel and T. Yanagawa. Incorporating historical controls in testing for a trend in proportions. *Journal of the American Statistical Association*, 81(396):1095–1099, 1986.

H. Jeffreys. Some tests of significance, treated by the theory of probability. 31(2): 203–222, 1935.

L. L. Kupper, C. Portier, M. D. Hogan, and E. Yamamoto. The impact of litter effects on dose-response modeling in teratology. *Biometrics*, 42(1):85–98, 1986.

L. G. León Novelo, A. Womack, H. Zhu, and X. Wu. A bayesian analysis of quantal bioassay experiments incorporating historical controls via bayes factors. *Statistics in Medicine*, 36(12):1907–1923, 2017.

S. D. Peddada, G. E. Dinse, and G. E. Kissling. Incorporating historical control data when comparing tumor incidence rates. *Journal of the American Statistical Association*, 102(480):1212–1220, 2007.

N. T. Program et al. Ntp technical report on the toxicology and carcinogenesis studies of benzophenone (cas no. 119-61-9) in f344/n rats and b6c3f1 mice (feed studies). Technical report, 2006.

R. Van de Schoot, D. Kaplan, J. Denissen, J. B. Asendorpf, F. J. Neyer, and M. A. van Aken. A gentle introduction to bayesian analysis: Applications to developmental research. *Child Development*, 85(3):842–860, 2014.