BAYESIAN METHODS FOR NON-STANDARD MISSING DATA PROBLEMS

BY JERRY Q. CHENG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

David Madigan and Minge Xie

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2010

ABSTRACT OF THE DISSERTATION

Bayesian Methods for Non-Standard Missing Data Problems

by Jerry Q. Cheng

Dissertation Director: David Madigan and Minge Xie

Missing data presents challenges to statistical analysis in many applications such as clinical trials, cluster detection, etc. This thesis analyzes and develops methodologies in some non-standard missing data problems.

We first consider non-ignorable drop-out in longitudinal clinical trials. Common simple approaches such as complete case analysis or last observation carried forward can lead to biased estimates and underestimation of uncertainty. We pursue a model-based approach in the context of Bayesian framework to provide more useful inferences.

Second, non-compliance is another way to deviate from pre-designed protocols. Traditional methods circumvent the issue with simplifying assumptions such as intention to treat. Consequently they might produce misleading results. We adopt a counterfactual approach, known as the Rubin Causal Model, essentially reducing the analysis to a missing data problem. We address the issue in particular when drop-out is also involved.

In relation to the first two research topics to provide better and more accurate assessment of a treatment or procedure, we develop a Bayesian sequential meta-analysis framework to aggregate results from all available studies. We conduct a case study and build a risk profile of a treatment to provide early alert of emerging problems.

Last, the question whether a spatial pattern is randomly distributed has been of interest in many applications. We extend and generalize a latent model approach to overlapping cluster detection. We employ this methodology to design an urban mobile sensor network for the surveillance of nuclear materials. With simulation studies, we demonstrate that the method is efficient and powerful in detection of overlapping clusters.

Acknowledgements

I am deeply thankful for my two advisers, Prof. David Madigan and Prof. Minge Xie. Their invaluable guidance and support over the years have helped me make progress in my research work. I am blessed to have the opportunity to work with them and learn from up close experiences their professionalisms and dedications to research and career advancement. I would like to thank my other dissertation committee members - Dr. Tamra Carpenter and Prof. Donald Hoover for their precious time and guidance on this thesis.

Next, I'd like to extend my appreciations to professors and researchers in Department of Statistics and DIMACS at Rutgers University for their wonderful lectures and advices. I would like to thank Prof. Fred Roberts for providing me an excellent opportunity to take on several interesting and important projects with DIMACS, which eventually led to a part of this thesis.

Further I would like to thank my fellow graduate students in Statistics Department who have helped me along the way with their friendships, stimulating discussions and idea exchanges.

This thesis will not be possible without the blessings from my family.

I am indebted to my parents - Jianmin Cheng and Jiakeng Zhou for their care and understanding, to my brother Minghe Cheng and his family for their support and taking up the role of primary care-taker of our aging and ailing parents.

Finally yet foremost, to my wife - Yingying Chen, my lovely son - Jeffrey Cheng, I must thank for giving me the love, encouragement, and patience to keep me going through the five plus wonderful years in my life.

Dedication

To my family on both sides of the Pacific Ocean.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

Missing data presents challenges to statistical analysis in many applications such as clinical trials, cluster detection, etc. This thesis analyzes and develops methodologies in some non-standard missing data problems. We present the main ideas of these methods below.

Non-Ignorable Drop-out

In longitudinal trials, especially those with serious adverse events, subjects can often drop out of the study and we are no longer able to collect their measurement data. In handling missing data due to drop-outs, various methods exist in the literature, for example, complete case analysis, observed data method, imputation, last observation carried forward, model-based method, etc. Simple approaches tend to yield biased estimates. Comparatively, the model-based method is very complex but it can explore various realistic assumptions about the drop-out process.

In this thesis, we apply the model-based method in the context of Bayesian framework in which missing data are treated the same way as parameters. We model response and drop-out status jointly. As a case study, we pick a large three-year Vioxx clinical trial. We divide the three year trial span into twelve 90-day periods and build longitudinal models for the endpoints of systolic blood pressure (SBP) and confirmed thrombotic cardiovascular (CVT) adverse event.

We find that the effect of the drug on SBP shows marked sensitivity to assumptions about the missing data mechanism, whereas the effect on CVT events shows less sensitivity.

Non-Compliance

Besides drop-out, non-compliance is another way to deviate from pre-designed protocols where a patient stops taking his assigned treatment or takes an incorrect dose. Traditional analysis methods such as intention-to-treat compare responses solely on the basis of treatment assignment, but not on the treatment that a subject actually receives. Consequently, these methods can produce misleading results when substantial non-compliance exists.

In dealing with this problem, we adopt a counterfactual approach, known as the Rubin Causal Model, essentially reducing the analysis to a missing data problem. The treatment effect is estimated as a population level causal effect for the compliers. We address the problem in particular when drop-out is also involved. We build various models on the counterfactual variables of response, compliance and missing status for each subject. These models allow explorations of the data and different drop-out mechanisms. Since they might yield different results, we use Bayes factors for model selection and calculate the weighted average of the causal effect from the model averaging perspective. All these steps constitute a general and flexible framework to deal with the important issue of non-compliance combined with drop-out in clinical trials.

As with the drop-out study, we use the same Vioxx trial to conduct the analysis using the same endpoints. We find that for both SBP and CVT, the causal effects of Vioxx are positive. This indicates that the drug increased patients' blood pressure and also the likelihood to incur a CVT adverse event.

Bayesian Sequential Meta-Analysis

Complex drugs usually involve multiple underlying mechanisms and can produce adverse effects on various timelines. As a result, major drug development processes typically include many different clinical trials focused on different doses, routes of administration, indications, endpoints, etc. In relation to the first two research topics to provide better and more accurate assessment of a new treatment, we develop a Bayesian sequential meta-analysis method to aggregate results from all available studies.

A flexible framework for analysis of randomized trials was proposed by Spiegelhalter *at el.* (1994). They used a Bayesian approach in which a family of priors was adopted instead of a single prior. In addition, a simple Gaussian model to summarize hazard ratio of interest was used in each individual trial. Our research extends this study to the meta-analysis context and generalizes it in two aspects. First, we model directly event counts with Poisson distributions in both treatment arms thereby reducing model assumptions. Second, we build hierarchical models to explore different data structures with varying complexities. Since studies are often put into groups with similar design and target, we introduce a block configuration in these models.

For a case study, we pick the set of placebo-controlled trials for Vioxx. We model the relative risk of the CVT adverse events using fixed effect, random effect, block effect and random block effect model under the family of priors over an eight year period. The result indicates that four years before the drug was withdrawn from the market, the probability of the relative risk exceeding 1.1 - a safety threshold, is above 50%. This demonstrates that Bayesian sequential meta-analysis can monitor risk profiles over the life time of a drug and provide earlier alert for emerging problems.

A Latent Model Approach for Cluster Detection

The question whether a spatial pattern is randomly distributed over space after adjusting for a known inhomogeneity has been of interest in many applications. A traditional statistical method to detect a cluster of events is via scan statistics which has been successful in detecting a single cluster and multiple clusters of fixed sizes. However, problems arise for detecting clusters of varying sizes.

To address this problem, Sun (2008) developed a latent model approach for multiple spatial cluster detection. With probability distributions to model the clusters and mimic the sample generation process, the approach used an EM/MCMC algorithm and likelihood inference to estimate model parameters, detect significant clusters, and identify their locations and sizes. In the study, all the clusters are assumed to be non-overlapping. This assumption simplifies the theoretical formation as well as its implementation. We extend and generalize the latent model approach to overlapping case. Simulation studies suggest that our method can achieve better detection powers than the original algorithm when the true clusters overlap.

In an application with potentially significant impact, we employ this methodology to design an urban mobile sensor network for the surveillance of nuclear materials. We design the network with a certain number of vehicles (e.g., taxicabs in New York City), on which nuclear sensors and Global Position System (GPS) tracking devices are installed. Real time readings of the sensors are to be processed at a central command center, where the latent modeling algorithm is used to analyze data and detect significant clusters which might indicate the locations of nuclear sources.

Thesis Organization

The rest of the thesis is organized as follows. Chapter 2 discusses our method to handle non-ignorable drop-out in longitudinal clinical trials. Chapter 3 deals with the issue of non-compliance using causal inference approach. Chapter 4 proposes a general and flexible framework to conduct Bayesian sequential meta-analysis. Chapter 5 extends and generalizes the latent model approach to overlapping cluster detection. Finally Chapter 6 concludes the thesis and provides future research directions.

Chapter 2

Non-Ignorable Drop-out in Longitudinal Clinical Trials

2.1 Introduction

Longitudinal clinical trials follow up participants over time to collect treatment-related measurements. Participants can deviate from the prescribed protocol in a variety of ways, either by ceasing to be evaluated (drop-out) or by failing to comply with their assigned treatments (non-compliance). We consider drop-out in this chapter and defer the non-compliance problem in the next one.

Participants that drop out provide no subsequent measurements thereby creating a missing data problem. The literature describes a wide variety of methods for handling missing data (Little and Rubin, 2002; Schafer, 1997). In many clinical trial settings, the standard methodology used to analyze incomplete data is based on such methods as *last observation carried forward* (LOCF), *complete case analysis* (CC), or simple forms of imputation. There are underlying assumptions for these methods. For example, the LOCF assumes that the missing data after the patient's withdrawal are the same as the last value observed for that patient. The CC methods discard all patients who did not complete their trials and analyze the remaining data. Imputation method fill in the missing data values using information from observed data. Therefore these simple approaches tend to oversimplify problems with strong and often unrealistic assumptions. As a result, they can lead to biased estimates and an underestimation of uncertainty.

Model-based approaches (Little, 1993; Carpenter *et al.*, 2002; Little and An,

2004) can present significant statistical and computational challenges, but can also provide more useful inferences. The general approach that we adopt here follows that of Carpenter *et al.* (2002).

The outline of the chapter is as follows. Section 2.2 describes Vioxx and its clinical trials that motivate our work. Section 2.3 details the elements of our approach to drop-out. Section 2.4 presents a simulation study to illustrate the drop-out mechanism and its impact on the estimation of treatment effect. Section 2.5 applies our approach to the Vioxx Data. Section 2.6 concludes the chapter with discussions.

2.2 Data

Vioxx is a COX-2 selective, non-steroidal anti-inflammatory drug (NSAID). The FDA approved Vioxx in May 1999 for the relief of the signs and symptoms of osteoarthritis, the management of acute pain in adults, and for the treatment of menstrual symptoms. The COX-2 class of drugs offered the hope of lower rates of gastrointestinal adverse effects as compared with standard NSAIDs like naproxen and ibuprofen. Instead, studies would eventually show that Vioxx causes an array of cardiovascular thrombotic side effects such as myocardial infarction, stroke, and unstable angina, leading to Merck's September 2004 withdrawal of the Vioxx from the market.

The APPROVe (Adenomatous Polyp Prevention on Vioxx) trial was the key study that lead to the withdrawal of Vioxx. APPROVe was a multi-center, randomized, placebo-controlled, double-blind study to determine the effect of 156 weeks (three years) of treatment with Vioxx (25mg daily) on the recurrence of neoplastic polyps of the large bowel. The trial enrolled 2,600 patients with a history of colorectal adenoma. We exclude 13 patients that inadvertently took the wrong dose of Vioxx. Of the remaining 2,587 patients 728 dropped out of the

study. Of these patients, 410 had been randomly assigned to Vioxx whereas 318 were assigned to placebo.

Upon enrollment, the trial gathered an array of demographic and medical data from each patient. Patients then paid office visits periodically, though not at rigidly fixed intervals. Those visits captured vital sign measurements, laboratory testing results, as well as information on potential adverse events. In what follows, and to highlight methodological nuances, we focus on two particular safety endpoints, systolic blood pressure (SBP) and confirmed thrombotic cardiovascular (CVT) adverse events.

The possibility that drop-out could be causally related to either or both of our safety endpoints seems well founded. Increased blood pressure might well lead a subject to drop out of the trial, as could the prequelae of a serious thrombotic event. If, as is now believed, Vioxx causes these events, the Vioxx arm of the trial could be deprived of the very patients most likely to suffer these adverse events. All published analyses of the APPROVe trial assume drop-out is unrelated to safety endpoints of interest ("ignorable") and hence could plausibly underestimate the effects of interest.

2.3 Drop-out in Clinical Trials

2.3.1 Notations

We consider a randomized longitudinal clinical trial including N patients over J time periods. For subject i at time period j, let $y_{i,j}$ denote primary response, $i = 1, \ldots, N$,$j = 1, \ldots, J$. In what follows, $y_{i,j}$ can be real valued - $y_{i,j} \in \Re$ (e.g., SBP), or binary - $y_{i,j} \in \{0, 1\}$ (e.g., CVT). Let Y denote the $N \times J$ matrix of primary responses for all the subjects at all time periods. Let $y_i = Y_{i,1:J}$ denote row i of Y - responses at all time periods for ith subject.

We assume that there are C covariates such as age, sex, diabetes status,

treatment type, etc. Let X denote the $N \times C$ covariate matrix for all subjects. $x_i = X_{i,1:C}$ denotes the covariates for subject i.

Let M denote the $N \times J$ matrix for the dropout status for all the subjects at all time periods. The (i, j) element of M takes a value of one when the subject i drops out at time period j, and zero otherwise. We will only consider monotone missingness in this chapter - once a patient drops out he cannot return to the trial. So, the row vector $m_i = M_{i,1:J}$ comprises a set of 0's possibly followed by a set of 1's. This vector represents the dropout status for subject i for period 1 through period J. We use the terms missingness and dropout interchangeably.

For simplicity in notation, we drop the index i and j when there is no confusion, and use only one covariate X though there might be several in practice. The primary response variable is split according to whether its value is recorded or missing with a shorthand notation: $y = \{y^{obs}, y^{miss}\}$. Equations (2.1) and (2.2) describe the relationship between Y and M for subject i:

$$y_i^{obs} = \{y_{i,j} : M_{i,j} = 0, j \in \{1, 2, \ldots, J\}\} \tag{2.1}$$

$$y_i^{miss} = \{y_{i,j} : M_{i,j} = 1, j \in \{1, 2, \ldots, J\}\} \tag{2.2}$$

Note that $|y_i^{obs}| + |y_i^{miss}| = J$. ($y^{obs}$, y^{miss}, M, X) represents the "complete data."

2.3.2 Joint Model for Response and Missingness

We start with the full likelihood function of the complete response y and M given covariates X. Let Ω denote the parameter space for the joint probability model. We factor Ω into subspaces that dissect the full likelihood function. Let $\Omega_R \subset \Omega$ relate the response y to the covariates X, $\Omega_{MAR} \subset \Omega$ relate M to y^{obs} and X, and $\Omega_{NIM} \subset \Omega$ relate M to y^{miss},

$$\Omega = \Omega_R \cup \Omega_{MAR} \cup \Omega_{NIM}. \tag{2.3}$$

We consider parameter vectors $\beta \in \Omega_R, \alpha \in \Omega_{MAR}$, and $\delta \in \Omega_{NIM}$ where β is of primary interest.

The *complete data likelihood function* is then:

$$p(y, M|X, \alpha, \beta, \delta) = p(y^{obs}, y^{miss}, M|X, \alpha, \beta, \delta) \tag{2.4}$$

Since we consider that subjects in the trial are independent, the joint likelihood in equation (2.4) factors as:

$$p(y^{obs}, y^{miss}, M|X, \alpha, \beta, \delta) = \prod_{1 \leq i \leq N} p(y_i^{obs}, y_i^{miss}, m_i|x_i, \alpha, \beta, \delta) \tag{2.5}$$

We can further factorize the ith component in Equation (2.5) into the marginal model of y_i given x_i and the conditional model of m_i given y_i and x_i:

$$p(y_i^{obs}, y_i^{miss}, m_i|x_i, \alpha, \beta, \delta) = p(m_i|y_i^{obs}, y_i^{miss}, x_i, \alpha, \delta)p(y_i^{obs}, y_i^{miss}|x_i, \beta) \tag{2.6}$$

Equation (2.6) represents a so-called *selection model*, in which dropout depends on both observed and unobserved response variables as well as covariates. The observed data likelihood is given by:

$$p(y^{obs}, M|X, \alpha, \beta, \delta) = \int p(y^{obs}, y^{miss}, M|X, \alpha, \beta, \delta) dy^{miss}. \tag{2.7}$$

Coupled with the selection model (2.6), this becomes:

$$p(y^{obs}, M|X, \alpha, \beta, \delta) = \prod_{1 \leq i \leq N} \int p(m_i|y_i^{obs}, y_i^{miss}, x_i, \alpha, \delta)p(y_i^{obs}, y_i^{miss}|x_i, \beta) dy_i^{miss}. \tag{2.8}$$

Before we discuss inference using (2.8), we first consider different standard simplifications of $p(m_i|y_i^{obs}, y_i^{miss}, x_i, \alpha, \delta)$. Little and Rubin (1987) considered three general categories for the *missing data mechanism*:

1. *Missing completely at random (MCAR)* where missingness does not depend on the response - either y^{obs} or y^{miss}:

$$p(M|y^{obs}, y^{miss}, X, \alpha, \delta) = p(M|X, \alpha) \tag{2.9}$$

2. *Missing at random (MAR)* where missingness does not depend on the unobserved response y^{miss}:

$$p(M|y^{obs}, y^{miss}, X, \alpha, \delta) = p(M|y^{obs}, X, \alpha) \tag{2.10}$$

3. *Non-ignorable (NIM)* where missingness depends on y^{miss} and possibly on y^{obs} and covariates.

In the NIM scenario, the parameter δ plays a key role, defining, as it does, the relationship between dropout and missing values. Unfortunately, the data provide little or no information about δ, and checking the sensitivity of the ultimate inferences to different choices of δ is central to our approach. In APPROVe, as we discussed above, an assumption of MAR or MCAR may not be tenable.

Inference Under *MCAR* and *MAR*

The assumption of *MCAR* simplifies the observed likelihood for ith subject using (2.9):

$$\begin{aligned} p(y_i^{obs}, m_i|x_i, \alpha, \beta, \delta) &= \int p(m_i|x_i, \alpha)p(y_i^{obs}, y_i^{miss}|x_i, \beta)dy_i^{miss} \\ &= p(m_i|x_i, \alpha) \int p(y_i^{obs}, y_i^{miss}|x_i, \beta)dy_i^{miss} \\ &= p(m_i|x_i, \alpha)p(y_i^{obs}|x_i, \beta) \end{aligned} \tag{2.11}$$

Similarly under *MAR* the observed likelihood becomes:

$$\begin{aligned} p(y_i^{obs}, m_i|x_i, \alpha, \beta, \delta) &= \int p(m_i|y_i^{obs}, x_i, \alpha)p(y_i^{obs}, y_i^{miss}|x_i, \beta)dy_i^{miss} \\ &= p(m_i|y_i^{obs}, x_i, \alpha)p(y_i^{obs}|x_i, \beta) \end{aligned} \tag{2.12}$$

Equations (2.11) and (2.12) show that the joint observed likelihood of response and missingess can factor into a product of the likelihood of the observed response and that of the missing data mechanism. We represent this partition in a shorthand notation as:

$$[y^{obs}, M|X, \alpha, \beta, \delta] = [M|y^{obs}, X, \alpha][y^{obs}|X, \beta] \tag{2.13}$$

If $\Omega_R \cap \Omega_{MAR} = 0$ and $\Omega_R \cap \Omega_{NIM} = 0$, in other words β is distinct from α and δ, inference about β depends only on the $p(y^{obs}|X, \beta)$. Hence aforementioned CC method is justified. However, the CC method is often inefficient because a substantial amount of valuable information has to be discarded. Multiple imputation (MI) is the main approach to handling missing data under the MCAR or MAR assumption (Rubin, 1987, 1996; Schafer, 1999). The key step in MI is to impute missing data from the predictive distribution $[y^{miss}|y^{obs}, X, M]$. The default method for that is Markov Chain Monte Carlo (MCMC) with initial values derived from an initial EM-run. A further, non-parametric alternative is offered by Rubin's approximate Bayesian bootstrap approach (Rubin, 1987).

Inference Under *NIM*

Under NIM, no simplification of the likelihood is possible and the integration in (2.8) cannot be accomplished in closed form. Therefore, we will use an approximation to this integral.

Computationally, the maximum likelihood estimate (MLE) for missing data can be derived using EM (Expectation and Maximization) (Dempster *et al.*, 1977) and extensions. However, when data are sparse or incomplete, likelihood surfaces can behave poorly and the EM algorithm may not converge to the MLE (Schafer 1997). This may occur with multivariate incomplete data where sample size is small, rate of missing data is high and models are over-parameterized.

Instead of the observed likelihood function (2.8), we can conduct analysis based on the full likelihood as in (2.6). MCMC methods are well suited in this context although they require the specification of prior distributions for all model parameters. Combining the full likelihood with these priors leads to a *full probability model* that can then be analyzed according to the Bayesian paradigm. The main point is that it is generally feasible to obtain a posterior sample from the model parameters and missing values $(\alpha, \beta, y^{miss})$, that is, a sample from the

distribution

$$[\alpha, \beta, y^{miss}|y^{obs}, M, X, \delta] \tag{2.14}$$

In the Bayesian framework, y^{miss} is unknown and is treated as a parameter. With (2.6) for $i = 1, ..., N$, and priors for the parameters, we can obtain all marginal and conditional distributions are well defined, as is the distribution in (2.14). For example, we are interested in estimating the treatment effect from $[\beta|y^{obs}, M, X, \delta]$. Clearly, this involves inestimable non-ignorable drop-out parameter δ whose behavior cannot be accessed directly from the data at hand. In a sensitivity study, we vary the value of δ and assess sensitivity to departures from ignorability (MCAR and MAR).

2.4 A Simulation Study

Before turning to the APPROVe data, we conduct a simulation study to shed some light on the robustness of the inferences about the primary parameters of interest to mismatches between the true value of δ and the assumed value of δ in a sensitivity analysis.

We set the number of subjects N to 800. For subject i, we generate data as follows:

- Step 1: With equal probability, generate covariate x_i to take value of 1 or 0, where 1 indicates the assignment of a hypothetical drug and 0 placebo.

- Step 2: Generate the response variable y_i via a simple regression model:

$$y_i = \beta x_i + \epsilon_i \tag{2.15}$$

  where β=2, $\epsilon_i \sim N(0, \sigma^2)$ and σ=1, $i = 1, ..., 800$.

- Step 3: Generate the missing data indicator m_i via a logistic regression model:

$$logit(P(m_i = 1)) = \alpha_0 + \alpha_1 x_i + \delta y_i \tag{2.16}$$

where $\alpha_0 = 1$, $\alpha_1 = 1$, δ varies from -1 to 1 with increment of 0.2.

- Step 4: With the missing data indicator m_i controlled by δ from the previous step, reset the response to be missing ($y_i = NA$) when $m_i = 1$.

For each preset value of δ, we generate a set of complete data: (y, x, m). We fit the data jointly with models (2.15) and (2.16) to estimate β, α_0 and α_1. In the drop-out model (2.16), we vary the assumed value of non-ignorable parameter (δ^* hereafter) from -1 to 1 with increment of 0.2.

With each combination of (δ,δ^*) and diffuse priors, we draw 10,000 MCMC iterations discarding first 200 to obtain the Bayesian estimate (posterior mean) of β - the "treatment effect" ($\hat{\beta}$ hereafter). The MCMC convergence appears satisfactory. After smoothing via a two dimensional LOESS algorithm (Cleveland *et al.*, 1988; Cleveland and Grosse, 1991), Figure (2.1) shows the the contour of $\hat{\beta}$ as a function of (δ, δ^*).

The plot shows a striking pattern: the upper left corner over-estimates the true value of 2 while the lower right corner under-estimates. Along the diagonal where δ^* is close to the true δ, the estimates are close to the true value of 2.

The treatment effect β represents the expected change of the response between the trial drug and the placebo. When the underlying non-ignorable drop-out parameter δ is positive, the high valued response is more likely to be missing. If we do not acknowledge this and estimate the treatment effect only using the observed data, the $\hat{\beta}$ will be smaller, i.e., biased downward. If we misjudge the drop-out mechanism to use negative values of δ^* in the modeling, we will impute responses with smaller values. As the result, we will underestimate the treatment effect - this is what happens in the lower right corner of the plot. Similarly, we will overestimate the treatment effect when the true δ is negative but we assume a positive δ^* as can be seen in the upper left corner of the plot.

From the plot, we can also observe that when the missing mechanism is indeed

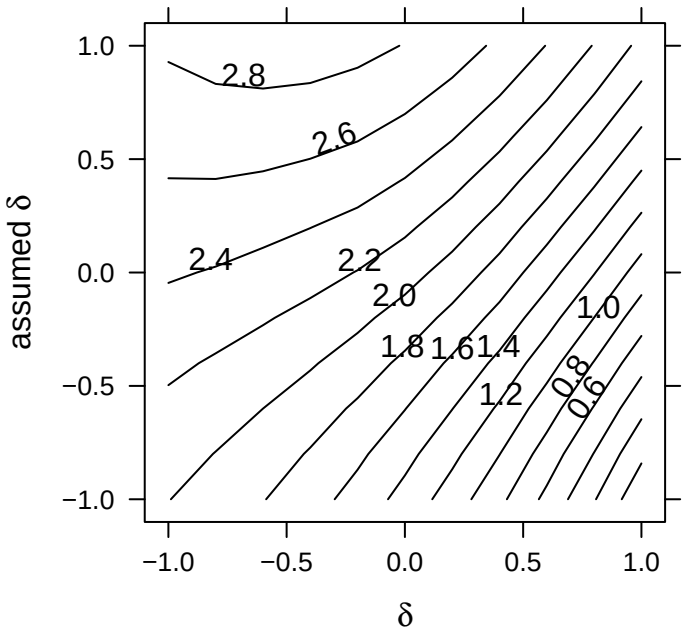


Figure 2.1: Smoothed contour plot of the estimated treatment effect. The horizontal axis is δ and the vertical axis is δ^*.

ignorable ($\delta = 0$), we will get biased estimates if we set δ^* to be either positive or negative for NIM. Meanwhile when the missing is truly NIM, the choice of δ^* impacts on the bias on the estimate.

This simulation shows that the estimation of the treatment effect can depend on the non-ignorable drop-out mechanism. With a range of δ^*, we can get an interval on which the estimated parameter falls. If the interval is narrow, the estimation is not sensitive to the choice of non-ignorable drop-out parameter.

2.5 Analysis of Vioxx Data

To facilitate the longitudinal study of the trial, we divide the three year trial span into twelve 90-day periods. The two endpoints have different rules to determine the drop-out status M. Since most patients did not re-enter the study once they

dropped out, we assume the monotone missingness pattern for the data analysis. The details to set the values for M are as follows:

- For SBP: find the last period which has the observed blood pressure for a subject. Then the periods after it will have $m = 1$ (missing), the earlier periods will have $m = 0$.
- For CVT: set $m = 1$ (missing) for periods after all patient contact ceases and set $m = 0$ for other periods.

The rules to set the response y are as follows:

- For SBP: in the quarters with $m = 0$, set $y = NA$ if there is no response; set the average to y if there are multiple measurements in the same period.
- For CVT: in the quarters with $m = 0$, set $y = 1$ if there is at least one CVT event; set $y = 0$ otherwise. The reason for the latter is that the missingness means the subject does not have any CVT events when the subject stays in the study.
- For periods with $m = 1$, set $y = NA$ for both endpoints.

From model selection procedure, we include two covariates namely treatment type (Tr) and an indicator of high cardiovascular disease risk (Hc). Both covariates are binary: Tr=1 for assignment of Vioxx, 0 for placebo; Hc=1 if the subject has high cardiovascular disease risk, 0 otherwise.

2.5.1 Non-Longitudinal Analysis

We begin with a non-longitudinal analysis focusing on the treatment effects at the end of the trial with the selection model (2.6). For subject i ($i = 1, ..., 2584$), the response model is:

- For SBP:

$$y_i = \beta_0 + \beta_{tr} Tr_i + \beta_h Hc_i + \epsilon_i \tag{2.17}$$

where $\epsilon_i \sim N(0, \sigma^2)$, y_i is the change of SBP from the baseline period to the period measured between 990 and 1080 days - a 90 day window before the end of the 3rd year - if such a measurement is available.

- For CVT:

$$logit(p(y_i = 1)) = \beta_0 + \beta_{tr} Tr_i + \beta_h Hc_i \tag{2.18}$$

where Y_i is the indicator from day 1 to day 1080.

The drop-out model for M is:

$$logit(p(m_i = 1)) = \alpha_0 + \alpha_{tr} Tr_i + \alpha_h Hc_i + \delta y_i \tag{2.19}$$

As in Section (2.3.2), we call δ in (2.19) the non-ignorable drop-out parameter since it relates the drop-out probability to the current response which might be missing. Hence we can make the missing data mechanism an $MCAR$ when δ=0, and an NIM when δ is not zero.

As in the simulation study, we use WinBUGS to perform $MCMC$ with 20,000 iterations discarding the first 1,000. MCMC convergence appears to be satisfactory. We list the posterior mean $\hat{\beta_{tr}}$ and its standard deviation in Table 2.1 for the two endpoints from various assumed values in δ.

Table 2.1: Bayesian estimate of treatment effects at the end of the third year under various assumed values of the non-ignorable drop-out parameter.

Scenario	SBP		CVT	
	Estimated Mean	Estimated SD	Estimated Mean	Estimated SD
δ=4	0.475	0.135	0.670	0.281
δ=0	0.300	0.124	0.669	0.280
δ=-4	0.213	0.126	0.669	0.280

From the table, the results under the $MCAR$ (δ=0) and NIM ($\delta \neq 0$) are quite different for SBP. This indicates that the estimated treatment effect is sensitive to the underlying non-ignorable missing data assumption. Moreover $\hat{\beta_{tr}}$ decreases as δ decreases. The explanation for this phenomenon follows. If we assume that a bigger increase in blood pressure is more likely to result in drop-out, we can undertake appropriate weighting to reduce unbiasedness, i.e., $\delta = 4$ in this case. Here we impute more higher values of SBP which are missing. Therefore the resulting treatment effect will be in the higher end. In the case of $\delta = -4$, the same reasoning works except that now we will compensate more in the lower blood pressure region. Thereby we obtain a smaller treatment effect estimate. For this endpoint, the result of estimation is sensitive to the choice of δ under NIM.

The result of $\hat{\beta_{tr}}$ for CVT follows the same trend. However, because the number of incidences for this endpoint is quite small so that the estimated standard error is large (about one half of $\hat{\beta_{tr}}$). Therefore the estimation is not sensitive to the assumption of NIM versus $MCAR$. The $MCAR$ with covariates will work fine for CVT.

2.5.2 Longitudinal Study

We expand the single time-point study in Section 2.5.1 to the longitudinal setting. The total number of periods is 12 and each period has 90 days.

We specify here the two likelihood functions in the selection model (2.6). The response model for patient i (i = 1, ..., 2584) at period j (j=1,...,12) is:

- For SBP:

$$y_{i,j} = \beta_j^0 + \beta_j^{tr} Tr_i + \beta_j^h Hc_i + \beta_j^{tm} j + \epsilon_{i,j} \tag{2.20}$$

  where $\epsilon_{i,j} \sim N(0, \sigma^2)$.

- For CVT:

$$logit(p(y_{i,j}=1)) = \beta_j^0 + \beta_j^{tr} Tr_i + \beta_j^h Hc_i + \beta_j^{tm} j \tag{2.21}$$

where β_j^{tm} is introduced for the possible time-dependence of y.

Because of the serious adverse events from Vioxx, we attribute responses such as blood pressure and CVT occurrence from current or past office visits to possible dropout from the trials. Hence we propose the model for drop-out status $m_{i,j}$ to include the end point observation from both current period and the previous period. Equation (2.22) shows the model for the first period:

$$logit(p(m_{i,1}=1)) = \alpha_1^0 + \alpha_1^{tr} Tr_i + \alpha_1^h Hc_i + \alpha_1^{tm} + \delta y_{i,1} \tag{2.22}$$

For the other periods (j=2,...,12), we add an additional term, i.e., $Y_{i,j-1}$ the response from the previous period :

$$logit(p(m_{i,j}=1)) = \alpha_j^0 + \alpha_j^{tr} Tr_i + \alpha_j^h Hc_i + \alpha_j^{tm} j + \alpha^{prev} y_{i,j-1} + \delta y_{i,j} \tag{2.23}$$

Again δ in (2.22) and (2.23) is the non-ignorable drop-out parameter. In (2.23) we make the missing data mechanism an $MCAR$ when δ=0 and α^{prev}=0, an MAR when both δ=0 and α^{prev} is not 0, and an NIM when δ is not zero.

Following the inference under NIM in Section (2.3.2), we combine the likelihood functions in Equations (2.20)-(2.23) with diffuse (but proper) prior distribution. For the priors of the same group of parameters for different time periods, we adopt a hierarchical approach under which the parameters come from the same hyper-prior distribution thereby smoothing the values over periods. For j=1,2....

12,

$$\begin{aligned}
\beta_j^0 &\sim N(\mu_\beta^0, (\sigma_\beta^0)^2) \\
\beta_j^{tr} &\sim N(\mu_\beta^{tr}, (\sigma_\beta^{tr})^2) \\
\beta_j^h &\sim N(\mu_\beta^h, (\sigma_\beta^h)^2) \\
\beta_j^{tm} &\sim N(\mu_\beta^{tm}, (\sigma_\beta^{tm})^2) \\
\alpha_j^0 &\sim N(\mu_\alpha^0, (\sigma_\alpha^0)^2) \\
\alpha_j^{tr} &\sim N(\mu_\alpha^{tr}, (\sigma_\alpha^{tr})^2) \\
\alpha_j^h &\sim N(\mu_\alpha^h, (\sigma_\alpha^h)^2) \\
\alpha_j^{tm} &\sim N(\mu_\alpha^{tm}, (\sigma_\alpha^{tm})^2)
\end{aligned}$$

With 200,000 iterations and 1,000 burn-ins, the $MCMC$ convergence appears satisfactory. We list the posterior mean $\widehat{\mu_\beta^{tr}}$ and its standard deviation in Table 2.2 for the two endpoints from various values in (δ, α^{prev}) pair.

As in the non-longitudinal setting, the effect of Vioxx on blood pressure shows marked sensitivity to assumptions about the missing data mechanism, whereas the effect on CVT events shows less sensitivity.

Table 2.2: Bayesian estimate and its deviations under various values of the non-ignorable drop-out parameter.

Scenario	SBP		CVT	
	Estimated Mean	Estimated SD	Estimated Mean	Estimated SD
$\delta = 0$, $\alpha^{prev} = 0$	0.23	0.05	0.49	0.30
δ=4	0.32	0.05	0.53	0.22
δ=0	0.23	0.05	0.52	0.30
δ=-4	0.18	0.06	0.48	0.30

2.6 Discussion and Conclusion

In this chapter, we explore a Bayesian approach to handle the missing data resulting from drop-out. We emphasize sensitivity to assumptions about the ignorability of the drop-out mechanism. We present a simulation study to demonstrate the robustness of the inferences about treatment effect to mismatches between true value of non-ignorable parameter and its assumed value. This exploration clearly provides insights to the non-ignorable drop-out issue. We also explore other simulations with a general form of response such as adding an intercept term, etc. Although we observe the impact of treatment effect estimation due to different assumed non-ignorable parameters, the patterns are not as obvious as those in Figure 2.1 and explanations are not as straightforward.

In the analysis of longitudinal clinical trials, we advocate a hierarchical prior distribution approach of same group of parameters from different time periods. By doing so, we smooth the variations from these parameters' prior distributions and reduce the standard deviations of the parameters' posterior estimates.

We apply the techniques to analyze specific cardiovascular endpoints in APPROVe, a placebo-controlled randomized clinical trial concerning the pain killer Vioxx. The data sets suggested that the treatment effect on the SBP change over baseline period is sensitive to the non-ignorable drop-out parameter. Hence we adopt NIM for the data analysis. Otherwise we will have biased results. But for CVT, we find it unnecessary to use NIM as the missing data mechanism.

Chapter 3

Non-Compliance and Complier Average Causal Effect

3.1 Introduction

Longitudinal clinical trials follow up with patients over a period of time to evaluate their responses to a particular treatment. Due to various reasons, especially when there exist serious side-effects from a new drug or treatment, patients often do not adhere to their trial protocols. There are two possible forms of such a deviation: *drop-out* and *non-compliance*. We considered and dealt with the drop-out issue in Chapter 2. Besides dropping out of trials, subjects may fail to comply with their assigned treatments ("non-compliance"). For example, a patient can stop taking the assigned drug, or takes the incorrect dose, or takes the incorrect drug (if he has the access to it). Meanwhile he might still stay in the study and supply measurement data.

The traditional and standard methods to evaluate drugs' effectiveness do not consider the complication of non-compliance. For example, intention-to-treat (ITT) studies a drug on basis of the assignment, instead of the actual treatment that the subject receives. Consequently, whenever there are a large number of non-compliances, ITT might produce misleading results.

As an alternative and complimentary tool, Rubin Causal Model combines Bayesian analysis with counterfactual concepts. In this chapter, we adopt this approach and essentially reduce the analysis to a missing data problem. We build elaborate models and provide an accurate estimate of treatment effect -

a population level causal effect for the compliers. We address the problem in particular when drop-out is also involved. We model jointly the variables of response, compliance and missing status for each subject. From this, we allow explorations of the data and various drop-out mechanisms. Since different models might yield different results, we use Bayes factors for model selection and calculate the weighted average of the causal effect from model averaging perspective. All these steps constitute a general and flexible framework to deal with the important issue of non-compliance coupled with drop-out in clinical trials.

The rest of the chapter is arranged as follows. In Section 3.2, we elaborate our Bayesian approach to handle both non-compliance and drop-out. Section 3.3 conducts a simulation study to illustrate the concepts. In Section 3.4, we conduct a case study of the same Vioxx trial as in Chapter 2 using one continuous and one binary endpoint. Section 3.5 concludes the chapter with discussions and future research directions.

3.2 Non-compliance and Drop-out

3.2.1 Notations and Rubin Causal Model

A statistical study for causal effects compares the results of two or more treatments on a population of units, each of which in principle could be exposed to any of the treatments (Rubin, 1990). In what follows, we shall assume that the trial comprises two treatments which we label "T" or "1" for an experimental new treatment ("T" for Treatment) and "C" or "1" for an existing or placebo therapy ("C" for Control). The trial follows N subjects for a specified time period and measures some health outcome Y (e.g. survival) at the end of that period. The idea of Rubin Causal Model (RCM) is to estimate the causal effect of T relative to C. Intuitively, this causal effect for a particular subject is the difference between the result if the subject had been exposed to T and the result if, instead, the

subject had been exposed to C (Rubin, 1978).

Let Y_i^j be the health outcome for subject i if *all* subjects were assigned to treatment j ($i = 1, ..., N; j = 0, 1$). We define the ITT causal effect of assignment for subject i to be $Y_i^1 - Y_i^0$. This definition does not make much sense without the Stable-Unit-Treatment-Value-Assumption (SUTVA) (Lewis, 1963; Rubin, 1978; Imbens and Rubin, 1997a,b). The assumption says that that Y_i^j is *stable* in the sense that it would take the same value for all other treatment allocations such as subject i receives treatment j. This assumption is not innocuous - the health outcome for subject A could depend on subject B's treatment assignment if, for example, A and B were in the same household and the treatment has a psychological component. However, SUTVA is generally not contentious in the randomized studies considered in this chapter. With SUTVA we can consider Y_i^j to be the outcome for subject i if subject i were assigned to treatment j. We note also that other causal effect definitions are possible, e.g. Y_i^1/Y_i^0. The difference seems to match the treatment effect concept well in the traditional clinical trial setting.

Population-level causal effects are usually of more interest than subject-level effects and we adopt the common approach of simply averaging the subject-level causal effects. In what follows we will be especially interested in sub-population average causal effect, such as : ave$(Y_i^1 - Y_i^0 | i$-th subject is male).

To help categorize population according to their compliance behavior, we define D_i^j to be the compliance indicator for the treatment that subject i would receive given the assignment j ($j = 0, 1$). For example, if a i-th subject receives the new treatment while he has been assigned with the control, then $D_i^0 = 1$. We now have a 4-vector of "semi-latent" variables for i-th subject: $(D_i^0, D_i^1, Y_i^0, Y_i^1)$. These variables are semi-latent in the sense that for any one subject, we will generally observe at most two of the four variables, i.e., either D_i^0 and Y_i^0, or D_i^1 and Y_i^1. For any particular subject, either or both potentially observable variables

sometimes termed as *potential outcomes* may be missing.

For each subject, $D_i = (D_i^0, D_i^1)$ describes the compliance behavior. Imbens and Rubin (1997a) distinguished four categories of subjects. For treatment j (j=0,1), subject i is a:

- *Complier* if $D_i^j = j$, or the subject complies to whatever he has been assigned to.
- *Never-taker* if $D_i^j = 0$, or the subject takes placebo all the time.
- *Always-taker* if $D_i^j = 1$, or the subject takes new treatment all the time.
- *Defier* if $D_i^j = 1 - j$, or the subject takes the opposite of what he has been assigned to.

With the above definition, we are ready to define some sub-population causal effects of interest. The most often used - the complier average causal effect (CACE) is give by:

$$CACE = avg_i(Y_i^1 - Y_i^0 | D_i^0 = 0, D_i^1 = 1). \tag{3.1}$$

Similarly we can define the defier average causal effect (DACE), the always-taker causal effect (AACE), and the never-taker causal effect (NACE). Of the four subpopulation causal effects, AACE and NACE do not address causal effects of the receipt of treatment since the former compares outcomes both with treatment, and the latter compares outcomes both without treatment. For compliers, assignment to treatment agrees with receipt of treatment and CACE compares outcomes with drug to outcomes without drug. For such complier subjects, following Imbens and Rubin (1997a), we will attribute the effect on Y of assignment to treatment to the effect of receipt of treatment. This attribution is what trialists typically do in randomized trials with full compliance. The DACE is also of some interest although in what follows, we will focus on the CACE as the primary estimand of interest.

3.2.2 Non-Compliance Coupled with Drop-out

When a subject stops to be evaluated, we say that he drops out of the study. In Chapter 2, we treated drop-out under a Bayesian framework and jointly modeled the response and missing data indicator. In practice, both non-compliance and drop-out can occur. Yau and Little (2001) developed a model for inference about the CACE which allows for the inclusion of baseline covariates and handles drop-outs in the repeated outcome measures. Their study used *missing at random* (MAR) as the missing data mechanism.

In this chapter, we generalize the problem dealing both non-compliance and drop-out with non-ignorable missing data mechanism. Similar to the definitions of Y_i^j and D_i^j, we extend the missing data indicator in the context of causal inference. We define M_i^j to be an indicator for the drop-out status that the subject i would have given the assignment $j(j = 0, 1)$. M_i^j is binary with the value of 1 for drop-out and 0 otherwise.

The set of the data for i-th subject now has six random variables: Y_i^0, M_i^0, D_i^0, Y_i^1, M_i^1 and D_i^1. When we know the treatment assignment for a subject, the variables in the other treatment group (also termed as *counter-factual group*) are all latent. Figure 3.1 lists various scenarios of a trial involving non-compliance and drop-out when the assigned treatment is a new drug. For example in Case 3, at the time point of consideration, we will have the measurement data for Y^1, D^1=1 (complier), M^1=0 (still in study), the variables in the counter factual group: Y^0, D^0, M^0 are all unknown (NA). With the assignment of a placebo, Case 1' to 5' are the counterparts of Case 1 to 5. The values for the six variables are set in Table 3.1 for all the cases.

After setting up the variables, we are ready to build models on them. Recent developments in Bayesian computation render the estimation of the CACE straightforward. Imbens and Rubin (1997a) present a detailed description of a particular approach to estimation. Here we frame the task in the context of

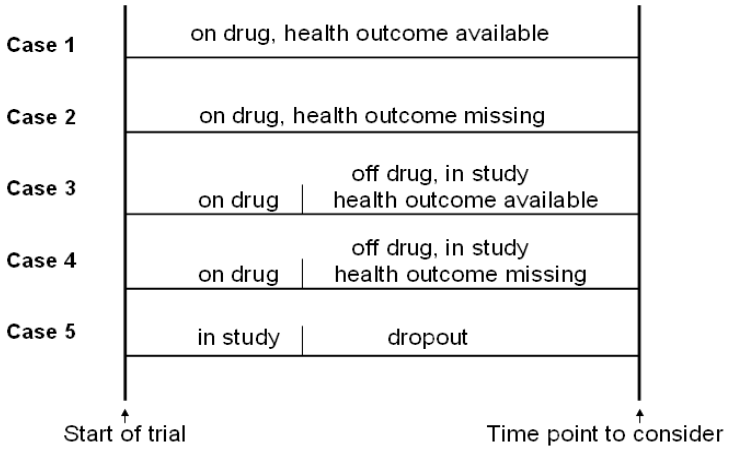


Figure 3.1: Different cases of a trial involving non-compliance and drop-out when the assigned treatment is a new drug.

Table 3.1: Various cases for the values of health outcome, compliance and missing indicators.

Case	Treatment Assignment	Y_i^1	D_i^1	M_i^1	Y_i^0	D_i^0	M_i^0
1	1	available	1	0	NA	NA	NA
2	1	NA	1	0	NA	NA	NA
3	1	available	0	0	NA	NA	NA
4	1	NA	0	0	NA	NA	NA
5	1	NA	NA	1	NA	NA	NA
1'	0	NA	NA	NA	available	1	0
2'	0	NA	NA	NA	NA	1	0
3'	0	NA	NA	NA	available	0	0
4'	0	NA	NA	NA	NA	0	0
5'	0	NA	NA	NA	NA	NA	1

Bayesian graphical models (Spiegelhalter and Lauritzen, 1990; Madigan and York, 1995) which simplifies the procedures and makes extensions to models involving covariates and multiple compliance indicators direct and transparent, at least in principle.

For example, we can build five possible Bayesian graphical models in Figure 3.2 for the six random variables (Y^0, M^0, D^0, Y^1, M^1, D^1). The covariates are not shown in the model for the sake of an easy presentation. The covariates might have links to all other stochastic nodes in the models.

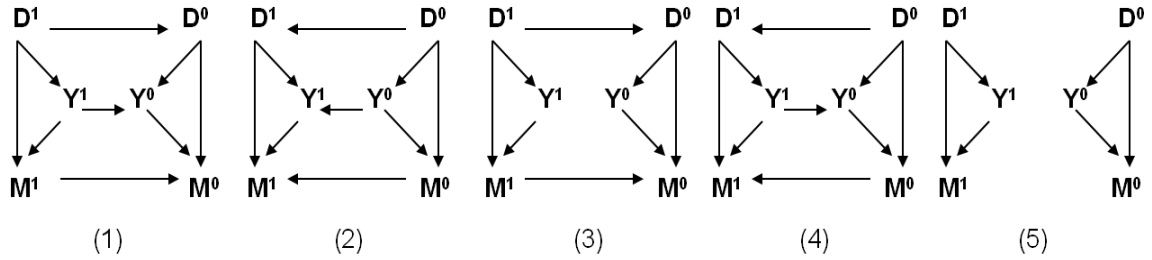


Figure 3.2: Bayesian network model: covariates are omitted

3.2.3 Bayesian Methods to Calculate CACE and Bayes Factor

With the aid of Bayesian graphical models, we can lay out the conditional likelihood functions for each node given its parent nodes including covariates. Using X as a covariate, we list the likelihood functions for Graphic Model 1 in Figure 3.2 for a i-th subject as follows:

1. Logistic regression models for conditional likelihood of binary compliance indicators: $[D_i^1|X_i]$ and $[D_i^0|D_i^1, X_i]$:

$$logit(p(D_i^1 = 1)) = \alpha_1^1 + \alpha_2^1 X_i$$
$$logit(p(D_i^0 = 1)) = \alpha_1^0 + \alpha_2^0 X_i + \alpha_3^0 D_i^1$$

2. Linear regression models for conditional likelihood of continuous health outcome, or logistic regression model for that of binary health outcome: $[Y_i^1|D_i^1, X_i]$ and $[Y_i^0|D_i^0, Y_i^1, X_i]$:

   - for continuous type

$$Y_i^1 = \beta_1^1 + \beta_2^1 X_i + \beta_3^1 D_i^1 + \epsilon_i^1$$
$$Y_i^0 = \beta_1^0 + \beta_2^0 X_i + \beta_3^0 D_i^0 + \beta_4^0 Y_i^1 + \epsilon_i^1$$

   where $\epsilon_i^1 \sim N(0, \sigma_v^2)$, $\epsilon_i^0 \sim N(0, \sigma_p^2)$.

- for binary type:

$$logit(p(Y_i^1 = 1)) = \beta_1^1 + \beta_2^1 X_i + \beta_3^1 D_i^1$$
$$logit(p(Y_i^0 = 1)) = \beta_1^0 + \beta_2^0 X_i + \beta_3^0 D_i^0 + \beta_4^0 Y_i^1$$

3. Logistic regression models for conditional likelihood of drop-out indicators: $[M_i^1|Y_i^1, D_i^1, X_i]$ and $[M_i^0|D_i^0, Y_i^0, M_i^1, X_i]$:

$$logit(p(M_i^1 = 1)) = \gamma_1^1 + \gamma_2^1 X_i + \gamma_3^1 D_i^1 + \delta^1 Y_i^1$$
$$logit(p(M_i^0 = 1)) = \gamma_1^0 + \gamma_2^0 X_i + \gamma_3^0 D_i^0 + \gamma_4^0 M_i^1 + \delta^0 Y_i^0$$

The CACE is calculated as the mean of the average casual effect of compliers out of each MCMC iteration (see Appendix A).

$$\widehat{CACE} = \frac{1}{\sum_{i=1}^N \left[\widehat{D_i^1}(1 - \widehat{D_i^0})\right]} \sum_{i=1}^N \left[(\widehat{Y_i^1} - \widehat{Y_i^0})\widehat{D_i^1}(1 - \widehat{D_i^0})\right] \tag{3.2}$$

With noninformative priors for all the parameters in the likelihood function(α_1^1, α_2^1, ..., γ_1^0, γ_1^0, γ_2^0, γ_3^0, γ_4^0 as well as latent random variables), we use posterior draws of the latent variables from MCMC iterations to get its Bayesian estimate (posterior mean).

Bayes Factor for Model Selection and Averaging

We follow the same steps to fit other models and obtain the estimations of the CACE. Since different models might yield different results, we use Bayes factor (Kass and Raftery, 1995) for model selection. The Bayes factor for model M_1 against another model M_0 given data $\mathbb{D}$ is the ratio of the two marginal likelihoods. The marginal likelihood is the probability of the data $\mathbb{D}$ given a model, and is obtained by averaging over the priors assigned to the parameters. Let Φ denote the set of parameters in the models. The Bayes factor of model 2 against 1 B_{12} is

$$B_{21} = \frac{[\mathbb{D}|M_2]}{[\mathbb{D}|M_1]} = \frac{\int [\mathbb{D}|\Phi, M_2][\Phi|M_2]d\Phi}{\int [\mathbb{D}|\Phi, M_1][\Phi|M_1]d\Phi} \tag{3.3}$$

The Bayes factor expresses the support given by the data for one or the other of the models, in a similar way to the conventional likelihood ratio. However, unlike classical significance procedures, the Bayes factor does not tend to reject the null hypothesis more frequently as sample sizes become large. Taking twice the log of the Bayes factor gives the same scale as the conventional deviance and likelihood ratio statistics. Approximate values for interpreting B_{21} and $2\log_e B_{21}$ are as in Table 3.2 (Jeffreys, 1961; Kass and Raftery, 1995).

Table 3.2: Guidelines for Bayes factors.

B_{21}	$2\log_e B_{21}$	Interpretation
Under 1	Negative	Support model 1
1 - 3	0 - 2	Weak support for model 2
3 - 20	2 - 6	Support for model 2
20 - 150	6 - 10	Strong evidence for model 2
Over 150	Over 10	Very strong support for model 2

With missing data, from either non-compliance or drop-out, we partition the complete data into the observed and the missing parts: $\mathbb{D} = (\mathbb{D}^{obs}, \mathbb{D}^{miss})$. The "observed " Bayes factor is

$$\begin{aligned} B_{21}^{obs} = \frac{[\mathbb{D}^{obs}|M_2]}{[\mathbb{D}^{obs}|M_1]} &= \int \frac{[\mathbb{D}^{obs}, \mathbb{D}^{miss}|M_2]}{[\mathbb{D}^{obs}|M_1]} d\mathbb{D}^{miss} \\ &= \int \frac{[\mathbb{D}^{obs}, \mathbb{D}^{miss}|M_2]}{[\mathbb{D}^{obs}, \mathbb{D}^{miss}|M_1]} \frac{[\mathbb{D}^{obs}, \mathbb{D}^{miss}|M_1]}{[\mathbb{D}^{obs}|M_1]} d\mathbb{D}^{miss} \\ &= \int B_{21} \times [\mathbb{D}^{miss}|\mathbb{D}^{obs}, M_1] d\mathbb{D}^{miss} \end{aligned} \tag{3.4}$$

Since the integral in (3.4) does not have closed form in most cases, we approximate it by the Monte Carlo simulations. With sample draws from distribution $[\mathbb{D}^{miss}|\mathbb{D}^{obs}, M_1]$ - the result obtained from the MCMC steps from calculating CACE, we calculate the average of B_{21} with $\mathbb{D}^{miss}$ being replaced by the samples. Several methods are available to calculate the marginal likelihood $[\mathbb{D}|M_i]$ (i=1,2). For binary response, we use the exact method from Cooper and Herskovits (1992)

and Heckerman (1996). For real valued response, we use asymptotic approximation such as Laplace's method or the Bayesian information criterion (BIC) method. The approximation methods are quite computationally intensive since they require maximization of a likelihood function in each MCMC step. Appendix B details some of the commonly used methods to compute Bayes factor.

Assume there are K available models. Let $CACE_i$ is the complier average causal effect from the i-th model, B_{i1}^{obs} be the observed Bayes factor of i-th model against the first one ($B_{11}^{obs} = 1$). We can get the weighted average of the CACE from model averaging perspective as:

$$CACE = \frac{\sum_{i=1}^{K} B_{i1}^{obs} \times CACE_i}{\sum_{i=1}^{K} B_{i1}^{obs}} \tag{3.5}$$

3.3 A Simulation Study

Before analyzing the real data set, we conduct a simulation study to shed some light on model dependent results of CACE. We start to generate data from Graphical Model 1, then fit the data using Model 1 through 5.

For $i = 1, 2, ..., 1000$, the sequential steps to simulate data is as follows:

1. Generate treatment Trt_i and covariate X_i as two independent Bernoulli(0.5) random variables.

2. Generate D_i^1 via $logit(p(D_i^1 = 1)) = 4 + X_i$

3. Generate D_i^0 via $logit(p(D_i^0 = 1)) = -4 + X_i + \alpha D_i^1$

4. Generate Y_i^1 via $Y_i^1 = 2 + X_i + D_i^1 + \epsilon_i^1$ where $\epsilon_i^1 \sim N(0, 0.01)$

5. Generate Y_i^0 via $Y_i^0 = 1 + X_i + D_i^0 - \alpha Y_i^1 + \epsilon_i^0$ where $\epsilon_i^0 \sim N(0, 0.01)$

6. Generate M_i^1 via $logit(p(M_i^1 = 1)) = \beta + X_i + \delta^1 Y_i^1$

7. Generate M_i^0 via $logit(p(M_i^0 = 1)) = \beta + X_i + D_i^0 + \alpha M_i^1 + \delta^0 Y_i^0$

8. When $Trt_i = 0$ (placebo), set NA to D_i^1, Y_i^1, and M_i^1; when $Trt_i = 1$ (treatment), set NA to D_i^0, Y_i^0, and M_i^0. This step is to set the couterfactual group.

9. When $M_i^1 = 1$ (missing), set NA to Y_i^1 and D_i^1; when $M_i^0 = 1$ (missing), set NA to Y_i^0 and D_i^0

where α is set to 5, β to -8, δ^1 and δ^0 to 2.

For each subject, we simulate the value of (D^1, D^0, Y^1, Y^0) from Step 1 through 7. From that, we calculate the average of causal effect $Y^1 - Y^0$ for all the compliers, which is 18.5. The true CACE can be computed as:

$$\begin{aligned} E(Y^1 - Y^0 | D^1 = 1, D^0 = 0) &= E[Y^1 - (1 + X - 5Y^1) | D^1 = 1, D^0 = 0)] \\ &= 3.5 - 1.5 + 5 \times 3.5 = 19.5 \end{aligned}$$

Step 8 and 9 let the missingness and non-compliance take effect. There are 29% drop-outs and 11% non-compliers. With the simulated data, Using the five different graphical models under three assumed non-ignorable drop-out parameter δ^1 and δ^0, we draw 2,000,000 MCMC iterations discarding first half to obtain the Bayesian estimate (posterior mean) of CACE. The MCMC convergence appears satisfactory. From the simulation results in Table 3.3, the estimated CACE's from

Table 3.3: Estimated CACE and its standard deviation from simulated data.

	BGM 1	BGM 2	BGM 3	BGM 4	BGM 5
Drop-out Mechanism	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)
δ^1=2, δ^0=2	19.04 (0.06)	17.83 (0.22)	18.02 (0.16)	18.01 (0.18)	18.00 (0.17)
δ^1=0,δ^0=0	18.46 (0.31)	17.91 (0.23)	17.97 (0.20)	17.82 (0.31)	17.88 (0.23)
δ^1=-2,δ^0=-2	18.32 (0.09)	18.00 (0.22)	17.95 (0.20)	17.87 (0.26)	17.92 (0.27)

Model 1 (the true model), are very close to the true value, while the other models yield significantly different values, especially at the case where the non-ignorable dropout parameters match the true ones. The assumed values for δ^0 and δ^1 do not affect the results much in the Model 2 to 5.

Next we draw 2,000,000 MCMC iterations to calculate Bayes factors for Model 1 against the rest under three sets of assumed non-ignorable dropout parameters. Since both Laplace and BIC methods are computationally intensive, it is not feasible to conduct such large amount of computations in a reasonable time, we choose to apply the exact method to the standardized data (subtracting mean and dividing standard deviation) which is convert to a binary (above 0 or else). This method produces results close to Laplace method, yet much faster. Discarding the first half iterations, we list the results in Table 3.4. We can observe that the

Table 3.4: Log Bayes factors from simulated data.

Drop-out Mechanism	BGM 1 vs 2	BGM 1 vs 3	BGM 1 vs 4	BGM 1 vs 5
δ^1=2, δ^0=2	3.90 (3.53)	0.05 (0.26)	8713.31 (58.23)	8738.03 (74.55)
δ^1=0, δ^0=0	2.67 (2.12)	0.01 (0.12)	8663.27 (90.82)	8881.11 (136.15)
δ^1=-2, δ^0=-2	1.41 (1.40)	0.02 (0.03)	9155.37 (51.62)	9076.76 (90.81)

true model - Model 1 is very strongly supported against Model 4 and 5, strongly supported against Model 2, and weakly supported against Model 3 according to the guidelines in Table 3.2.

3.4 Analysis of Vioxx Data

Similar to Chapter 2, we use APPROVe study from Vioxx trials and choose two types of endpoints in this study: systolic blood pressure (SBP) which is continuous and confirmed thrombotic cardiovascular AE indicator (CVT) which is binary. Using the similar steps, we set values for response and missing data indicator at the end of the three year trial. We use the days in study along with days in treatment to assign the values for compliance indicator to assign values to the variable set $(Y_i^0, M_i^0, D_i^0, Y_i^1, M_i^1, D_i^1)$ for a i-th subject in the same spirit of the toy example of Section 3.2.2. With 2,000,000 iterations and 1,000,000 burn-ins, the $MCMC$ convergence appears satisfactory. We list the posterior

mean $CACE$ and its standard deviation in Table 3.5 for the three endpoints from various assumed values in (δ^v, δ^p) pair, under the five Bayesian graphical models.

Table 3.5: Estimated one-time point CACE for endpoints of SBP and CVT under five Bayesian graphical models (BGM).

Endpoint - SBP

	BGM 1	BGM 2	BGM 3	BGM 4	BGM 5
Drop-out Mechanism	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)
δ^v=4,δ^p=4	0.134 (0.040)	0.163 (0.035)	0.132 (0.033)	0.156 (0.032)	0.175 (0.043)
δ^v=0,δ^p=0	0.139 (0.038)	0.144 (0.037)	0.143 (0.039)	0.134 (0.036)	0.155 (0.035)
δ^v=-4,δ^p=-4	0.137 (0.037)	0.125 (0.039)	0.143 (0.039)	0.128 (0.038)	0.130 (0.040)

Endpoint - CVT

	BGM 1	BGM 2	BGM 3	BGM 4	BGM 5
Drop-out Mechanism	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)	$\widehat{CACE}$ (s.d.)
δ^v=4,δ^p=4	0.0111 (0.0052)	0.0125 (0.006)	0.0122 (0.005)	0.0108 (0.005)	0.0113 (0.005)
δ^v=0,δ^p=0	0.0108 (0.0052)	0.0118 (0.006)	0.0122 (0.006)	0.0103 (0.005)	0.0105 (0.005)
δ^v=-4,δ^p=-4	0.0109 (0.0051)	0.0115 (0.006)	0.0143 (0.006)	0.0108 (0.006)	0.0102 (0.005)

Table 3.6: (Log) Bayes Factors under five Bayesian graphical models (BGM) for SBP and CVT.

Endpoint - SBP

Drop-out Mechanism	BGM 1 vs 2	BGM 1 vs 3	BGM 1 vs 4	BGM 1 vs 5
δ^v=4,δ^p=4	-16.03 (8.85)	12.52 (1.66)	23422.86 (63.81)	23719.84 (48.31)
δ^v=0,δ^p=0	-17.95 (11.94)	13.17 (1.78)	23468.18 (58.12)	23693.45 (59.34)
δ^v=-4,δ^p=-4	-18.90 (9.57)	12.88 (2.00)	23868.18 (74.68)	23293.28 (80.34)

Endpoint - CVT

Drop-out Mechanism	BGM 1 vs 2	BGM 1 vs 3	BGM 1 vs 4	BGM 1 vs 5
δ^v=4,δ^p=4	-0.53 (2.02)	3.61 (5.66)	3.32 (2.90)	584.72 (29.07)
δ^v=0,δ^p=0	-0.68 (2.41)	3.28 (5.36)	3.59 (3.13)	599.62 (25.92)
δ^v=-4,δ^p=-4	-0.52 (2.53)	4.29 (1.86)	2.71 (2.75)	591.02 (31.45)

We observe that for SBP and CVT, the causal effects of Vioxx are all positive with significance. This indicates that the drug increased patients' blood pressure, the likelihood to incur a CVT adverse event. Also estimated CACE's are not sensitive to the different drop-out mechanisms.

Similar to the simulation study, we draw 2,000,000 MCMC iterations with

first half as burn-ins to calculate Bayes factors of Model 1 versus the rest under three sets of non-ignorable dropout parameters. The Bayes factor is computed from the exact method. The results are listed in Table 3.6. Model 2 is supported weakly against Model 1 for CVT and very strongly for SBP. Model 1 is strongly favored against Model 3 to 5 with SBP consistently offering stronger support than CVT.

3.5 Discussions and Conclusion

In the chapter, we propose a Bayesian framework to handle non-compliance at the presence of drop-out. This framework is generic and can handle non-ignorable missing data mechanism, accommodate various sensible models, assess model uncertainty from computing Bayes factor, and eventually obtain the weighted average CACE from model averaging perspectively.

Because of the drop-out and counter-factual data, we treat more than half of the data in the analysis as missing. As a result, our Bayesian analysis encounters significant computational challenges, which we address via a wide variety of MCMC and computational methods. For example, we apply block level move in Metropolis random walk to improve the sampling efficiency. Since the Bayes factor is extremely computationally intensive for the continuous response, we use binary approximation and achieve similar results to the Laplace and the BIC methods, and make the computation much faster. We have programmed the MCMC steps in R, WinBUGS, and C to take advantages of each computer language. We also tune the trial step parameters in Metropolis Hastings algorithm to increase its efficiency for different models and dropout mechanisms.

In principle, the study can be extended to longitudinal setting straightforward in theory (Appendix C). However the additional computational challenges need to be addressed.

Chapter 4
Bayesian Sequential Meta-Analysis

4.1 Introduction

Drug safety is an extremely important issue to consider in the drug development process. Complex drugs involve multiple underlying mechanisms and can produce adverse effects on various timelines. One key reason drugs may be used for years by millions of patients before risks of a drug become evident is that we do not have an active drug-surveillance system. As a result, a leading concern for the Food and Drug Administration (FDA) is to protect the public from drugs' risks as effective as possible (McClellan, 2007). A key incident leading to this concern was the 2004 withdrawal by Merck of rofecoxib (Vioxx) because of an apparent increased risk of serious cardiovascular events. Questions have been raised whether and when we can conduct analyses with all available information in order to provide early alerts for emerging problems. In the case of Vioxx, it would have been desirable to reveal cardiovascular risks before the drug's withdrawal. In light of this, there represents an opportunity to implement a set of systematic and cohesive processes to improve drug safety and effective use. Ideally this system will consist of data collections, case reporting and effective data analyses.

Usually major drug development processes include many different clinical trials focused on different doses, routes of administration, indications, endpoints, etc. To obtain a "big picture" result, we need to aggregate the results from these trials. Several statistical methodologies are relevant to handle this problem. Meta-analysis, sequential analysis, and Bayesian statistics are among them.

Meta-analysis is defined as a quantitative review and synthesis of results from related but independent studies (Normand, 1999). The objectives of a meta-analysis can be several-fold. By combining information over different studies, an integrated analysis will have more statistical power to detect a treatment effect than an analysis based on one study. When several studies have conflicting conclusions, a meta-analysis can be used to estimate an average effect or to identify a subset of studies associated with a beneficial effect. The basic meta-analysis method is a weighted average of point estimates (one from each study), with weights based on the standard errors of the estimates (Whitehead *at el.*, 1991). Commonly combination of estimates is achieved using either a fixed effect or a random-effects meta-analysis (DerSimonian *at el.*, 1986). Alternatives and extensions are frequently used. Sutton *at el.* (2007) conducts a detailed review of recent developments in meta-analysis.

In practice, meta-analyses are widely carried out to quantify the effectiveness of healthcare interventions (Sutton *at el.*, 2007). For example, the Cochrane Collaboration endeavors to collate and synthesize high-quality evidence on the effects of important healthcare interventions for a worldwide, multi-disciplinary audience, and publishes these in the Cochrane Database of Systematic Reviews (Cochrane, 2007).

In some situations, meta-analysis can be performed repeatedly whenever a new trial becomes available for inclusion. Termed as *sequential* or cumulative meta-analysis, it can retrospectively identify the point in time when a treatment effect first reached conventional levels of significance. As a result, sequential analyses can be used to decide whether enough evidence has been gathered in completed trials to make further trials unnecessary. Lau *et al.* (1992) used this approach to analyze 33 trials evaluating a new thrombolytic therapy for acute infarction from 1959 to 1988. They discovered a consistent, statistically significant reduction in total mortality was achieved in 1973 after only eight trials. Other references can

be found in Henderson *et al.* (1995) and Bollen *at al.* (2006).

In the meanwhile, Bayesian methods are becoming more frequently used in a number of areas of healthcare research, including meta-analysis. Though much of this increase has been directly as a result of advances in computational methods, it has also been partly due to their appealing nature. Bayesian tools offer great flexibility in meta-analyses to encourage a model-based approach to the combination of information from multiple sources. Therefore we can easily incorporate the "background" information pertinent to a certain clinical question being addressed. Similarly, in estimating the true effects of individual studies we can in some way "borrow strength" from other similar studies (Higgins *et al.*, 1996). Furthermore, Bayesian methods offer a unified modeling framework which overcomes issues such as the appropriate treatment of small trials, and a flexibility which allows the approach to be extended to consider distributions other than Gaussian for random effects, or to adjust for covariates through regression models (Warn *et al.* 2002). Because the posterior distribution is produced by simulation, via Markov Chain Monte Carlo (MCMC) techniques, inference regarding non-standard functions of the parameters is possible.

Bayesian analysis normally faces difficulty in choosing a suitable prior distribution. Some have suggested that several priors (reflecting clinicians' prior opinions) could be used (from skeptical prior to enthusiastic prior) to better reflect differing opinions about the likely benefit of the new treatment. A flexible framework for the analysis of randomized trials was proposed by Spiegelhalter *at el.* (1994). They used a Bayesian approach in which a family of priors was adopted instead of a single prior. In addition, a simple Gaussian model to summarize hazard ratio of interest was used in each individual trial.

This chapter proposes a Bayesian framework conducting meta-analysis sequentially in which information regarding treatment effects is updated as more trials become available. It provides the history of the evolution of the posterior

probability distribution of the clinical trial result and allows us to quantify the changes in our belief about the treatment effect as the data accumulate. The current posterior distribution becomes the new prior distribution at the next time instance. We extend the study by Spiegelhater *at el.* (1994) to a meta-analysis context and generalize the approach in several aspects. First, in addition to the normal approximation for the summary relative risk of an adverse event (AE) in a clinical trial, we model directly the event counts with Poisson distributions for the rate of the AE in each treatment arm. With that, we reduce model assumptions. Second, we build hierarchical models to explore different data structures with varying complexities. Since studies are often put into groups with similar design and targets, we introduce a block configuration in these models. For a case study, we pick the set of placebo-controlled trials for Vioxx. We focus on the relative risk (RR) of the cardiovascular thrombotic (CVT) adverse event using various models under the family of priors over an eight year period.

The rest of this chapter is organized as follows. Section 4.2 details our framework in dealing with sequential aggregation of individual trials. Section 4.3 applies the framework to a set of real clinical trials. We conclude the chapter with discussions in Section 4.4.

4.2 Bayesian Sequential Meta-analysis

4.2.1 General Setup and Notations

We consider placebo-controlled clinical trials and evaluate the RR in logarithmic scale of an AE from a new drug or treatment. Trials often have different beginning and ending times. We start at a time when a first trial is completed and estimate the RR. Following this, every time when a trial ends, we calculate the RR with the additional data. As a result, we can obtain the RR profile of a new drug sequentially in time. This profile can assist in the risk assessment of the new

drug. Depending on the models and assumptions, we can carry out the process by updating from addition data, or simply calculate the RR using all the available data up to one particular time.

Assume at any time t, there are n_t completed trials. Let n_i^C and n_i^T be the number of a particular AE event in the control arm and treatment arm respectively for the i-th trial. Let r_i be the estimate of the RR of the AE event under treatment arm versus placebo arm, δ_i be the true RR for the i-th trial, δ be the aggregated or average relative risk for all the trials. We calculate r_i using median-unbiased estimates (Rothman *et al.*, 2008). The standard deviation of $\log r_i$ is estimated by $\sqrt{1/n_i^C + 1/n_i^T}$. Some trial might have zero event count in either or both arms because the trial's size is small, or the AE is rare. We use a continuity correction by splitting count of 1 proportional to the subject number in each arm and add the value to the event count. Sweeting *et al.* (2004) has detailed discussions about this issue.

Sometimes trials might have the same design and target the same set of patients, the results of them can be reasonably assume to follow a same or related distribution model. We use the concept of block to present the group effect of the trials. We expand the notations for an i-th trial within a j-th study block. We use $r_{i,j}$ to denote the estimate of the RR, $\delta_{i,j}$ the true RR, $n_{i,j}^T$ and $n_{i,j}^C$ the event counts under treatment and control arm respectively, and δ_j the true RR for the j-th block.

Particularly interested in estimating δ via Bayesian hierarchical models, in what follows, we will explore multiple models suitable for different situations. Since this may results in different results from various models, we need a measurement for model selection. Spiegelhalter *et al.*, 2002 introduced the deviance information criterion (DIC) to compare competing models:

$$DIC = \overline{D(\theta)} + p_D \tag{4.1}$$

where $\overline{D(\theta)}$ is the average deviance (i.e. minus twice the log-likelihood) with respect to the posterior distribution of the parameter, θ, and p_D is the effective number of parameters for the model. The smaller the DIC is, the better its associated model is. In this study, the DIC is computed with the complete data (when all the trial data are available) for all the models.

4.2.2 Family of Priors

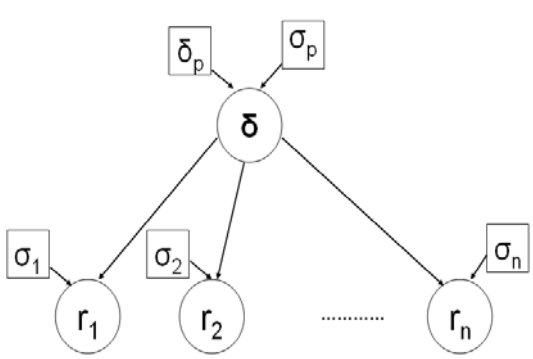
In our Bayesian meta-analysis, rather than focusing on single prior, we present analyses with a "family of priors" following Spiegelhalter *et al.* (1994). We propose a "skeptical prior" that represents *a priori* skepticism that the drug causes a certain adverse event. We also propose a "cautious prior" that represents a prior belief that the AE is not so implausible. We define two thresholds to help specify these priors. The upper threshold, δ_U, is the relative risk above which most reasonable people would agree that the drug should not be on the market. In this study, we choose $\delta_U = 1.75$. This reflects the belief that anything close to a doubling of AE risk would certainly be unacceptable. The lower threshold, δ_L, is roughly the value below which reasonable people would agree that it might be appropriate to have the drug on the market, and where informed patients could consider the risks and benefits on an individual basis. We chose $\delta_L = 1.1$. In between these two values, reasonable people could disagree.

We then construct Gaussian prior distributions around these thresholds. Specifically, the skeptical prior is centered on zero with just a 5% prior probability that the true hazard ratio δ exceeds δ_U (i.e., skeptical that the drug is dangerous). The cautious prior is centered on δ_U with a 5% probability that δ is less than δ_L. This is cautious insofar as it reflects a prior 50% probability that the true drug hazard ratio exceeds δ_U.

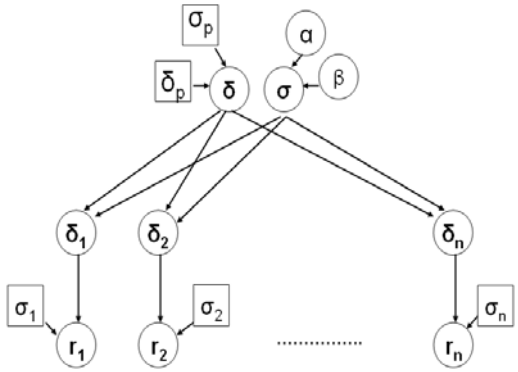
4.2.3 Gaussian Data Models

Following Spielgelhalter *et al.* (1994), we start with a simple Gaussian data model for the RR, thereby circumventing within-trial analytical issues. We build different models on the data: fixed effect, random effect, block effect and random block effect model with increasing complexities. Fixed effect Gaussian model is the simplest and analytically tractable. The other models do not have closed forms for the posterior distribution of the RR. So we resort to MCMC and aggregate the trials with all the available data up to that time.

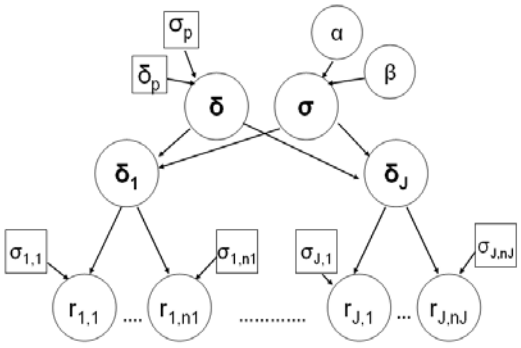
Figure 4.1 represents the four models in the form of Bayesian graphical model, which is a multivariate probabilistic model that uses a graph to represent a set of conditional independences (Madigan 2005).



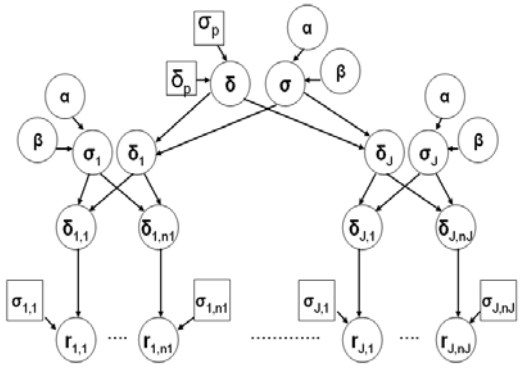
(1) Fixed Effect Model



(2) Random Effect Model



(3) Block Effect Model



(4) Random Block Effect Model

Figure 4.1: Various Bayesian graphical models under Gaussian data assumption.

Fixed Effect Model

In this part of the model, we assume that each clinical trial follows the same distribution with the population mean of δ and all the individual studies are estimating it. For a completed i-th trial at time t, we model its RR with a Gaussian distribution as follows:

$$\begin{aligned} r_i|\delta &\propto N(\delta, \sigma_i^2) \\ \delta &\propto N(\delta_p, \sigma_p^2) \end{aligned} \tag{4.2}$$

where σ_i^2 is the variance, δ_p and σ_p^2 are the mean and variance for the prior distribution of δ. Using a family of priors, we select the values of (δ_p,σ_p^2) from skeptical, cautious, reference, outrageous10 and outrageous1 priors as (0, 0.116), (0.560, 0.116), (0, 40.0), (0, 0.0055) and (0, 0.0579) respectively.

This fixed effect model has the advantage of simplicity and can produce the posterior distribution of relative risk analytically. As a result, we can easily update the posterior distribution of δ when the data from a newly completed trial is available. Without loss of generality, we assume that all n trials end at different times, T_1, T_2, ...,T_n. The relative risks for the trials are r_1, r_2, ..., and r_n.

1. Before time T_1, we have prior distribution $[\delta] \propto N(\delta_p, \sigma_p^2)$, which is one of the family of priors.

2. At time T_1, $[\delta|r_1] = \frac{[r_1|\delta][\delta]}{[r_1]} \propto [r_1|\delta][\delta] \propto N(\mu_1 = \frac{\frac{r_1}{\sigma_1^2}+\frac{\delta_p}{\sigma_p^2}}{1/\sigma_1^2+1/\sigma_p^2}, \tau_1^2 = \frac{1}{1/\sigma_1^2+1/\sigma_p^2})$, where μ_1 and τ_1^2 are posterior mean and variance respectively.

3. At time T_2, $[\delta|r_1, r_2] \propto [r_2|\delta, r_1][\delta|r_1][r_1] \propto [r_2|\delta][\delta|r_1]$ (because r_1 and r_2 are conditionally independent given δ) $\propto N(\mu_2, \tau_2)$, where $\mu_2 = \frac{\frac{r_2}{\sigma_2^2}+\frac{\mu_1}{\tau_1^2}}{1/\sigma_2^2+1/\tau_1^2}, \tau_2 = \frac{1}{1/\sigma_2^2+1/\tau_1^2})$

4. At time T_3, $[\delta|r_1, r_2, r_3] \propto [r_3|\delta, r_1, r_2][\delta|r_1, r_2] \propto [r_3|\delta][\delta|r_1, r_2] \propto N(\mu_3, \tau_3)$, where $\mu_3 = \frac{\frac{r_3}{\sigma_3^2}+\frac{\mu_2}{\tau_2^2}}{1/\sigma_3^2+1/\tau_2^2}, \tau_3 = \frac{1}{1/\sigma_3^2+1/\tau_2^2})$

5. In general, at time T_n, $[\delta|r_1, ..., r_n] \propto [r_n|\delta][\delta|r_1, ..., r_{n-1}] \propto N(\mu_n, \tau_n)$, where $\mu_n = \frac{\frac{r_n}{\sigma_n^2} + \frac{\mu_{n-1}}{\tau_{n-1}^2}}{1/\sigma_n^2 + 1/\tau_{n-1}^2}, \tau_n = \frac{1}{1/\sigma_n^2 + 1/\tau_{n-1}^2})$

Therefore, the posterior distribution of δ given the RR's $r_1, ...r_n$ can be equivalently obtained by sequentially updating the posterior mean and variance from Study 1 to n. When there are multiple trials ending at the same time instance, we can use the same procedure by treating these trials ending at slight different times since the aggregating result does not depend on the sequence of the trials.

Random Effect Model

Heterogeneity may exist for a variety of reasons: there may be differences in execution of the trials or in patient populations, or the trials may investigate different but still related treatments. Random effect model assumes that each individual study estimates its own, unknown, true effect which in turn is a perturbation about an overall population effect. In our study, we allow the true RR's to vary across studies and they follow a common probability distribution with mean of δ:

$$\begin{aligned} r_i|\delta_i &\propto N(\delta_i, \sigma_i^2) \\ \delta_i|\delta, \sigma^2 &\propto N(\delta, \sigma^2) \\ \sigma^2 &\propto Gamma(\alpha, \beta) \\ \delta &\propto N(\delta_p, \sigma_p^2) \end{aligned} \tag{4.3}$$

In this model, a Gamma prior is put on the variance parameter σ. The hyperparameter σ^2 follows a gamma distribution with fixed parameters α and β: $f(x; \alpha, \beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$. To make this distribution a non-informative one, normally we by default set $\alpha = 0.001$ and $\beta = 1000$ so that the mean of σ^2 is 1 and its variance is 1000. The parameters δ_p and σ_p^2 for the prior distribution of δ are chosen the same way as in (4.2).

In contrast to the fixed effect model, the extra layer of model on individual δ_i makes the posterior distribution $[\delta|r_1, ..., r_n]$ have inexplicit format (Appendix D). So we will rely on MCMC and use WinBUGS to get the posterior draws and estimate the posterior probability.

Block Effect Model

In this part of model, we have trials with several similar designs and subjects which we describe as study blocks. Within each block, we assume a fixed effect model so that all the trials follow the same distribution. The block level parameters, especially the true RR from each block, follow a common Gaussian distribution with the aggregated δ as its mean.

$$\begin{aligned} r_{i,j}|\delta_j &\propto N(\delta_j, \sigma^2_{i,j}) \\ \delta_j|\delta, \sigma^2 &\propto N(\delta, \sigma^2) \\ \sigma^2 &\propto Gamma(\alpha, \beta) \\ \delta &\propto N(\delta_p, \sigma^2_p) \end{aligned} \tag{4.4}$$

The prior distributions are chosen the same way for σ^2 and δ as in (4.3).

Random Block Effect Model

This is a model combining block effect with random effect models. Similar to block effect model, we can group studies into similar blocks while the group effects represented by δ_j for all the groups are assumed to come from a common normal distribution with the aggregated relative risk δ as the mean. However,

within each study block, we allow the trials to follow their own distributions:

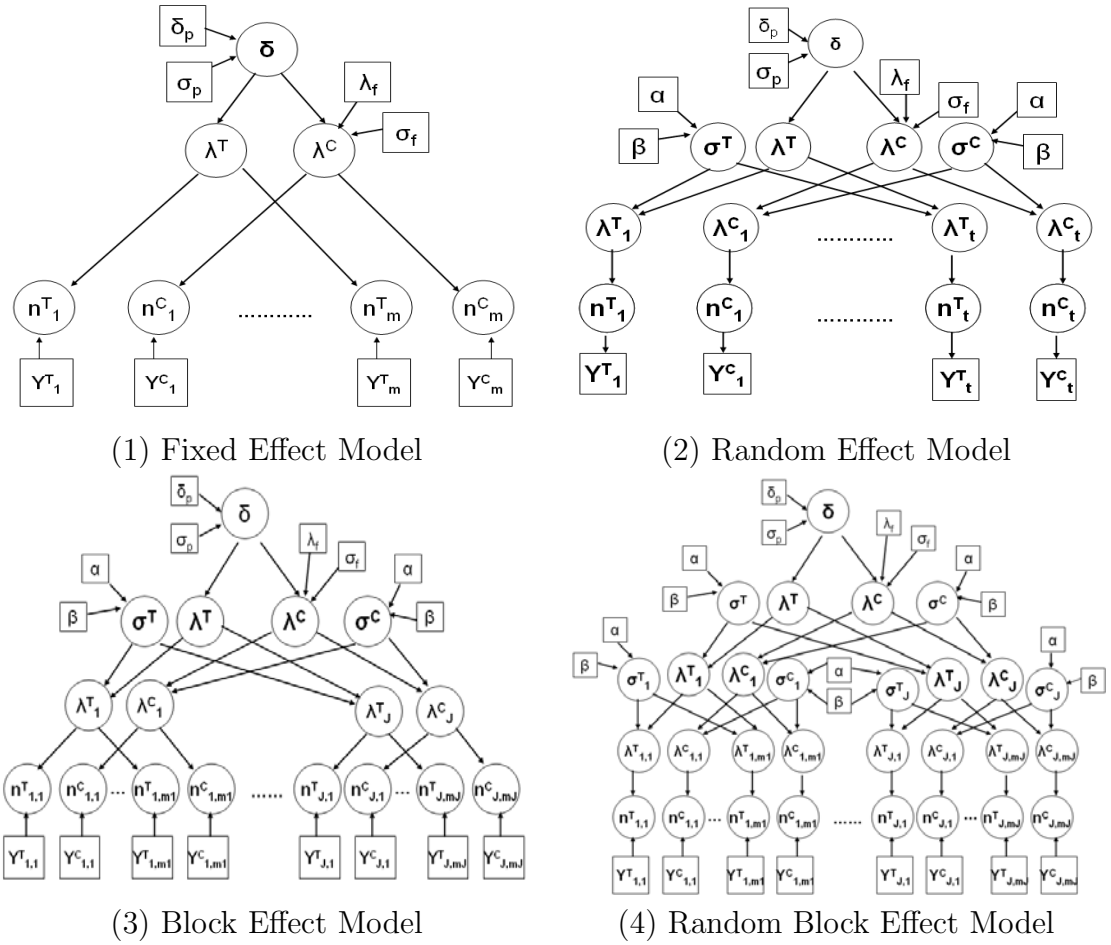
$$\begin{aligned} r_{i,j}|\delta_j &\propto N(\delta_{i,j}, \sigma^2_{i,j}) \\ \delta_{i,j}|\delta_j, \sigma^2_j &\propto N(\delta_j, \sigma^2_j) \\ \delta_j|\delta, \sigma^2 &\propto N(\delta, \sigma^2) \\ \sigma^2_j &\propto Gamma(\alpha, \beta) \\ \sigma^2 &\propto Gamma(\alpha, \beta) \\ \delta &\propto N(\delta_p, \sigma^2_p) \end{aligned} \tag{4.5}$$

The prior distributions are chosen the same way for σ^2_j, σ^2 and δ as in (4.3).

4.2.4 Poisson Data Models

The Gaussian data model in Section 4.2.3 greatly simplifies the analysis and Spiegelhalter *et al.* (1994) argues that in many situations the approximation is justified. Nonetheless, a more sophisticated approach that accounts for study-specific data model would be more satisfactory (Cheng and Madigan, 2010). In light of this, we model directly the counts of an AE event from both treatment arms using Poisson models since we have the information of patient year for each trial. Let λ^T_i and λ^C_i represent the Poisson rates, Y^T_i and Y^C_i the person year for the treatment arm and control arm respectively in i-th trail, λ^T and λ^C the common Poisson rate for all the trials under treatment arm and control arm respectively.

Similar to Section 4.2.3, we have four models for the approach. Figure 4.2 presents the graphical representations of the models. The parameter of interest - RR in logarithmic scale δ=log(λ^T/λ^C), is the top level note in these Bayesian graphical models.



(1) Fixed Effect Model (2) Random Effect Model

(3) Block Effect Model (4) Random Block Effect Model

Figure 4.2: Various models under Poisson data assumption.

Fixed Effect Model

Assume all the trials follow common distributions for the event counts under the two arms separately. The Poisson distributions have means of $\lambda^T Y_i^T$ and $\lambda^C Y_i^C$. We put a flat prior on the Poisson rate $\log \lambda^C$ and pick a set from the family of

priors on $\log \lambda^T$ conditioning on λ^C.

$$\begin{aligned} n_i^T|\lambda_i^T &\propto Poisson(\lambda^T Y_i^T) \\ n_i^C|\lambda_i^C &\propto Poisson(\lambda^C Y_i^C) \\ \log \lambda^C &\propto Normal(\lambda_f, \sigma_f^2) \\ \log \lambda^T|\lambda^C &\propto Normal(\delta_p + \log \lambda^C, \sigma_p^2) \\ \delta &= \log(\lambda^T/\lambda^C) \end{aligned} \tag{4.6}$$

where λ_f and σ_f^2 are mean and variance for the prior distribution of $\log \lambda^C$, δ_p and σ_p^2 are chosen to be the same as (4.2). Usually $\lambda_f = 0$ and $\sigma_f^2 = 10^4$ makes a non-informative prior distribution for $\log \lambda^C$.

Random Effect Model

In the same spirit of Section 4.2.3, each event count follows its own distinct distribution.

$$\begin{aligned} n_i^T|\lambda_i^T &\propto Poisson(\lambda_i^T Y_i^T) \\ n_i^C|\lambda_i^C &\propto Poisson(\lambda_i^C Y_i^C) \\ \log \lambda_i^T|\lambda^T &\propto Normal\left(\log \lambda^T, (\sigma^T)^2\right) \\ \log \lambda_i^C|\lambda^C &\propto Normal\left(\log \lambda^C, (\sigma^C)^2\right) \\ (\sigma^T)^2 &\propto Gamma(\alpha, \beta) \\ (\sigma^C)^2 &\propto Gamma(\alpha, \beta) \\ \log \lambda^C &\propto Normal(\lambda_f, \sigma_f^2) \\ \log \lambda^T|\lambda^C &\propto Normal(\delta_p + \log \lambda^C, \sigma_p^2) \\ \delta &= \log(\lambda^T/\lambda^C) \end{aligned} \tag{4.7}$$

where α and β follow (4.3), while λ_f, σ_p^2, δ_p and σ_p^2 are chosen to be the same as (4.6).

Block Effect Model

As a counterpart under Gaussian data model, we introduce study blocks to represent the similarities among subsets of the trials. We use λ_j^T and λ_j^C to represent the parameters for the jth study block, which the ith trial belongs to.

$$\begin{aligned}
n_{i,j}^T|\lambda_j^T &\propto Poisson(\lambda_j^T Y_{i,j}^T) \\
n_{i,j}^C|\lambda_j^C &\propto Poisson(\lambda_j^C Y_{i,j}^C) \\
\log \lambda_j^T|\lambda^T &\propto Normal\left(\log \lambda^T, (\sigma^T)^2\right) \\
\log \lambda_j^C|\lambda^C &\propto Normal\left(\log \lambda^C, (\sigma^C)^2\right) \\
(\sigma^T)^2 &\propto Gamma(\alpha, \beta) \\
(\sigma^C)^2 &\propto Gamma(\alpha, \beta) \\
\log \lambda^C &\propto Normal(\lambda_f, \sigma_f^2) \\
\log \lambda^T|\lambda^C &\propto Normal(\delta_p + \log \lambda^C, \sigma_p^2) \\
\delta &= \log(\lambda^T/\lambda^C)
\end{aligned} \tag{4.8}$$

where α, β, λ_f, σ_f^2, δ_p and σ_p^2 are chosen to be the same as (4.7).

Random Block Effect Model

As a counterpart under Gaussian data model, we have the combination of random effect models for the AE event counts of the treatment arms within each study

block as follows:

$$
\begin{aligned}
n_{i,j}^T|\lambda_j^T &\propto Poisson(\lambda_{i,j}^T Y_{i,j}^T) \\
n_{i,j}^C|\lambda_j^C &\propto Poisson(\lambda_{i,j}^C Y_{i,j}^C) \\
\log \lambda_{i,j}^T|\lambda_j^T &\propto Normal\left(\log \lambda_j^T, (\sigma_j^T)^2\right) \\
\log \lambda_{i,j}^C|\lambda_j^C &\propto Normal\left(\log \lambda_j^C, (\sigma_j^C)^2\right) \\
\sigma_j^T &\propto Gamma(\alpha, \beta) \\
\sigma_j^C &\propto Gamma(\alpha, \beta) \\
\log \lambda_j^T|\lambda^T &\propto Normal\left(\log \lambda^T, (\sigma^T)^2\right) \\
\log \lambda_j^C|\lambda^C &\propto Normal\left(\log \lambda^C, (\sigma^C)^2\right) \\
\sigma^T &\propto Gamma(\alpha, \beta) \\
\sigma^C &\propto Gamma(\alpha, \beta) \\
\log \lambda^C &\propto Normal(\lambda_f, \sigma_f^2) \\
\log \lambda^T|\lambda^C &\propto Normal(\delta_p + \log \lambda^C, \sigma_p^2) \\
\delta &= \log(\lambda^T/\lambda^C)
\end{aligned} \tag{4.9}
$$

where α, β, λ_f, σ_f^2, δ_p and σ_p^2 are chosen to be the same as (4.7).

4.3 Sequential Meta-Analysis in Clinical Trials Related to Vioxx

4.3.1 Vioxx and its Trials

Several plausible mechanisms exist by which Vioxx could cause a variety of cardiovascular thrombotic adverse events[1]. Some mechanisms operate on a short

[1]The specific events are acute myocardial infarction, unstable angina pectoris, sudden and/or unexplained death, resuscitated cardiac arrest, cardiac thrombus, pulmonary embolism, peripheral arterial thrombosis, peripheral venous thrombosis, ischemic cerebrovascular stroke, stroke (unknown mechanism), cerebrovascular venous thrombosis, and transient ischemic attack.

time scale while others may concern permanent damage that could lead to adverse events after Vioxx treatment has concluded (Antman *et al.*, 2007). Hence a complete characterization of Vioxx's cardiovascular risk profile should include an array of thrombotic events, on- and off-drug.

Except for a collection of smaller and shorter trials, and two larger trials that were terminated early when Vioxx was withdrawn from the market, Table E.1 in Appendix E provides a listing of the placebo-controlled trials that Merck conducted. The table lists placebo-controlled trails for Vioxx, which enrolled patients with rheumatoid arthritis ("RA"), with osteoarthritis ("OA") and with Alzheimer's disease ("ALZ"). The earliest of these trials concluded in 1996 while the latest concluded in 2004. Various publications reported the results of some of these trials individually and several meta-analysis of subsets of the trials exist (e.g., Konstam *et al.*, 2001; Reicin *et al.*, 2002; Weir *et al.*, 2003). However, as the trials progressed, no "big picture" analyses were performed. We contend that large-scale drug development programs such as that associated with Vioxx, should monitor safety on a sequential basis as data accumulate, taking into account all available data.

4.3.2 Analysis Results under Gaussian Data Models

With each model in (4.2), (4.3), (4.4) and (4.5), we conduct Bayesian computations and obtain the posterior drawings of the RR's of CVT on the individual trial level, study block level if relevant, and the aggregated level. The computation is repeated by using different prior distribution for δ from the family of priors. The posterior means and standard errors for the RR's are listed in Table E.2 of Appendix E. We summarize the results in Table 4.1. From the two tables, we can observe: 1). the aggregated RR's are similar under the same prior distribution for each of the four models; 2). the aggregated RR's differ significantly among choices of priors; 3). the block level RR's do not differ significantly between the

block effect model and the random block effect model.

Table 4.1: Estimated relative risks using Gaussian data assumption under different Bayesian graphical models and priors.

Aggregated-level RR

	Family of Priors				
Model	Skeptical	Cautious	Reference	Outrageous 10	Outrageous 1
Fixed Effect	0.33(0.10)	0.39(0.10)	0.37(0.11)	0.11(0.01)	0.30(0.10)
Random Effect	0.29(0.15)	0.40(0.14)	0.36(0.18)	0.05(0.07)	0.25(0.15)
Block Effect	0.30(0.14)	0.38(0.13)	0.35(0.15)	0.06(0.07)	0.26(0.13)
Random Block Effect	0.17(0.24)	0.42(0.23)	0.32(0.38)	0.02(0.07)	0.12(0.19)

Block-level RR

		Family of Priors				
Block	Model	Skeptical	Cautious	Reference	Outrageous 10	Outrageous 1
RA	Block Effect	0.31(0.24)	0.40(0.23)	0.37(0.25)	0.13(0.31)	0.28(0.24)
	Random Block	0.25(0.45)	0.43(0.44)	0.36(0.50)	0.13(0.44)	0.21(0.44)
OA	Block Effect	0.25(0.22)	0.33(0.21)	0.30(0.23)	0.03(0.24)	0.21(0.22)
	Random Block	0.13(0.35)	0.25(0.35)	0.19(0.37)	0.03(0.33)	0.10(0.34)
ALZ	Block Effect	0.34(0.13)	0.38(0.12)	0.37(0.13)	0.26(0.15)	0.32(0.13)
	Random Block	0.25(0.31)	0.35(0.30)	0.31(0.32)	0.18(0.31)	0.23(0.31)

Table 4.2: DIC for the four models at different priors using Gaussian data model.

	The Family of Priors				
Model	Skeptical	Cautious	Reference	Outrageous 10	Outrageous 1
Fixed Effect Model	54.2	54.1	54.3	54.4	54.3
Block Effect Model	58.9	58.7	59.0	60.8	59.0
Random Effect Model	59.0	58.9	59.1	59.1	61.2
Random Block Model	64.3	64.4	64.6	64.4	64.3

Next we use DIC as a criteria to select a model from the four. The DIC is computed with the complete data (when all the trial data are available) for the four models under the five different priors. From Table 4.2, the simplest fixed effect model has the smallest DIC value. This indicates that using Gaussian data assumption, the more complicated models are not necessary while the fixed effect model can effectively represent the data.

Last, focusing the fixed effect model, we look a closer look at the posterior distributions of the aggregated RR δ. We first draw the box plots of the posterior draws of RR's in logarithm scale at various time points and prior distributions in

Figure 4.3. We observe that 1). RR's under reference, skeptical and outrageous1 prior display the upward trend in time; 2). RR's under all priors are converging to be positive (in logarithmic scale).

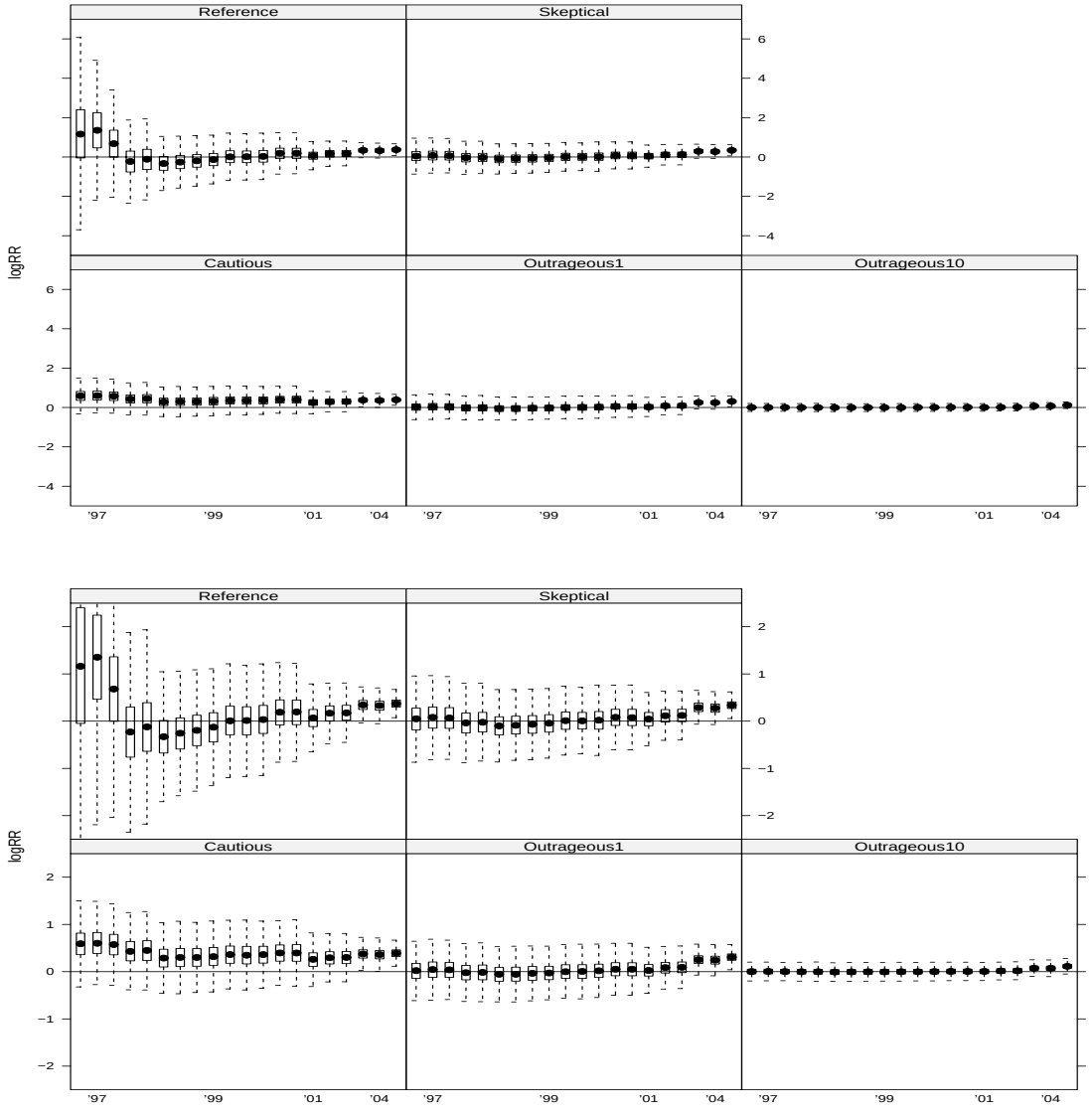


Figure 4.3: Box plots for the posterior draws of δ using Gaussian data model.

Now using the posterior draws from the fixed effect model, we estimate the probability of the (log)RR. Figure 4.4 shows the prior and posterior probability that $\delta > \delta_L$ under skeptical and cautious priors as well as three other priors.

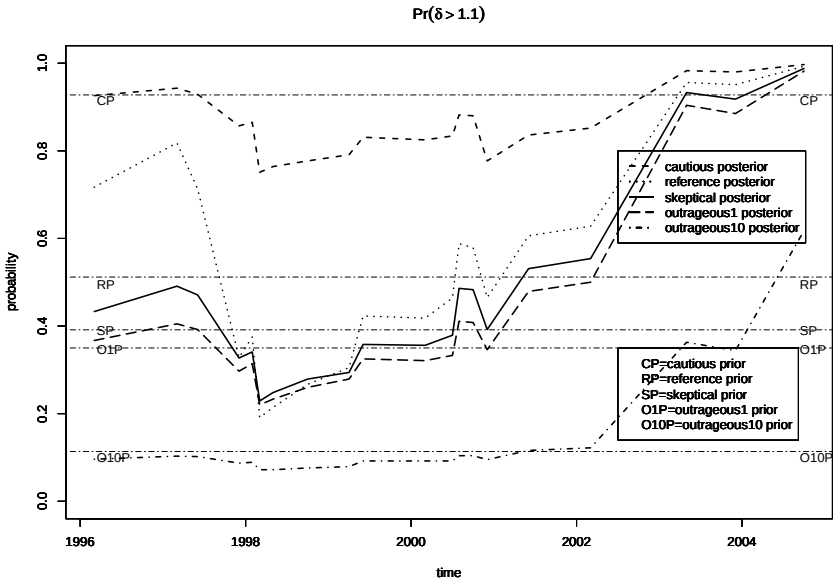


Figure 4.4: Sequential prior and posterior $\Pr(\delta > \delta_L)$ for all Vioxx placebo-controlled trials under Gaussian fixed effect model.

We observe that in the middle of year 2000, some four years before Vioxx was withdrawn from the market, the posterior probability that the relative risk was greater than δ_L exceeded 50% under all priors except the "outrageous10" prior. In our view, the two outrageous priors represent unreasonable prior beliefs in light of the available evidence at the outset. A striking feature of this analysis is the convergence toward $Pr(\delta > \delta_L) = 1$ under four of the five priors. This result is similar to Cheng and Madigan (2009) which computed the posterior probabilities every six months.

Figure 4.5 shows the posterior probability that $\delta > \delta_U$ over time. Clearly the data from the placebo-controlled trials provide minimal support for a value of δ as large as 1.75. Again, while the different priors provide different posterior estimates at the outset, they have largely converged by 2004.

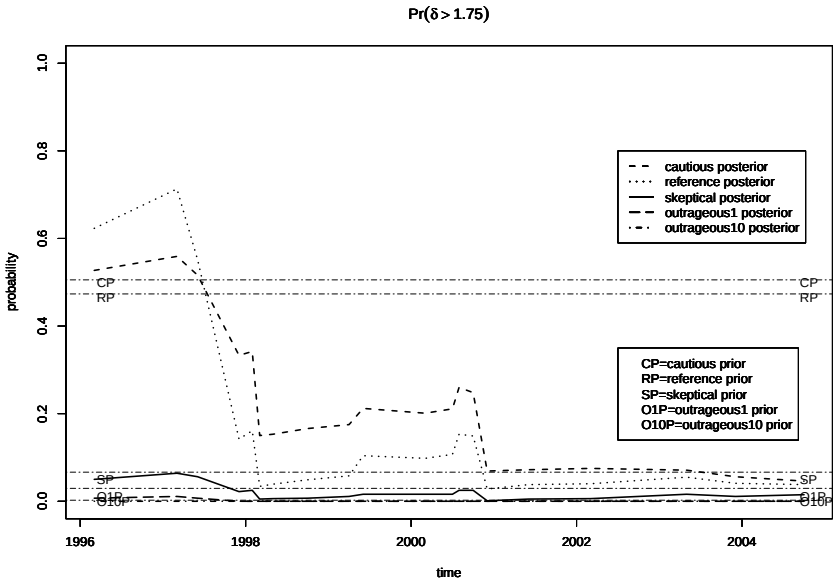


Figure 4.5: Sequential prior and posterior $\Pr(\delta > \delta_U)$ for all Vioxx placebo-controlled trials under Gaussian fixed effect model.

4.3.3 Analysis Results under Poisson Data Models

Before we apply the Poisson data assumption, we check the overdispersion from the data. With the available event count data, we estimate the mean and variance under both Vioxx and placebo arms in the study blocks of RA, OA and ALZ in Table 4.3. It appears that the event counts in the blocks of RA and OA follow Poisson distribution well, while the ones in ALZ block do not. This indicates the differences exit among blocks of studies and advocates the adoption of random effect models.

Table 4.3: Overdispersion check for Poisson models.

	Vioxx		Placebo	
Block	sample mean	sample variance	sample mean	sample variance
RA	0.031	0.035	0.015	0.014
OA	0.030	0.040	0.019	0.021
ALZ	0.046	0.427	0.034	0.145

Similar to Section 4.3.2, we conduct Bayesian computations and obtain the

posterior estimates of the relative risks. To balance the numerical instability and desirable non-informative prior distribution, we set $\lambda_f = 0$, $\sigma_f = 100$, $\alpha = 1$, and $\beta = 0.01$ in Models (4.6), (4.7), (4.8) and (4.9). So the mean and variance for $\log \lambda^C$ are 0 and 100, those for $(\sigma^C)^2$ (as well as $(\sigma^T)^2$) are 100 and 100,00.

Table 4.4: Estimated relative risks using Poisson data models under different Bayesian graphical models and priors.

Aggregated-level RR

		Family of Priors				
	Model	P1	P2	P3	P4	P5
	Fixed Effect	0.33(0.10)	0.39(0.10)	0.37(0.11)	0.11(0.06)	0.30(0.10)
	Random Effect	0.24(0.23)	0.49(0.24)	0.44(0.34)	0.02(0.08)	0.16(0.19)
	Block Effect	0.10(0.31)	0.54(0.29)	0.50(0.76)	0.01(0.07)	0.06(0.22)
	Random Block Effect	0.06(0.31)	0.56(0.33)	0.50(0.89)	-0.00(0.07)	0.04(0.23)

Block-level RR

		Family of Priors				
Block	Model	Skeptical	Cautious	Reference	Outrageous 10	Outrageous 1
RA	Block Effect	0.70(0.75)	0.84(0.74)	0.93(0.78)	0.74(0.75)	0.81(0.73)
	Random Block	0.72(1.01)	0.96(0.96)	0.99(1.08)	0.69(0.99)	0.66(0.98)
OA	Block Effect	0.57(0.43)	0.64(0.42)	0.64(0.44)	0.58(0.43)	0.61(0.43)
	Random Block	0.63(0.63)	0.77(0.64)	0.77(0.66)	0.66(0.64)	0.69(0.61)
ALZ	Block Effect	0.38(0.15)	0.37(0.15)	0.38(0.15)	0.38(0.16)	0.37(0.15)
	Random Block	0.19(0.69)	0.35(0.69)	0.34(0.77)	0.20(0.70)	0.20(0.71)

Table 4.5: DIC for the four models at different priors under Poisson models.

	The Family of Priors				
Model	Skeptical	Cautious	Reference	Outrageous 10	Outrageous 1
Fixed Effect Model	248.8	249.0	249.2	254.5	249.3
Block Effect Model	141.3	141.1	141.3	141.8	141.1
Random Effect Model	143.4	142.8	143.0	143.7	143.1
Random Effect with Block Model	141.7	141.8	141.8	141.7	141.8

The estimated relative risks at relevant levels from four models under families of priors are listed in Table E.3 of Appendix E. The parameter of interest - RR in logarithmic scale δ is $\log \lambda^T/\lambda^C$ as the top level note in the Bayesian Graphical Models. We summarize the results in Table 4.4: 1). the aggregated RR's are similar for block effect and randomized block effect model, similar for fixed effect model and random effect model; 2). RR's differ significantly between models with

blocks and one without blocks.

Next we want to select a best model from the four. The DIC is computed with the complete data for the four models at different priors. From Table 4.5, the block effect model is the best one. It also indicates that we are better off to use model with blocks than the simple fixed effect model which now has the biggest value in DIC.

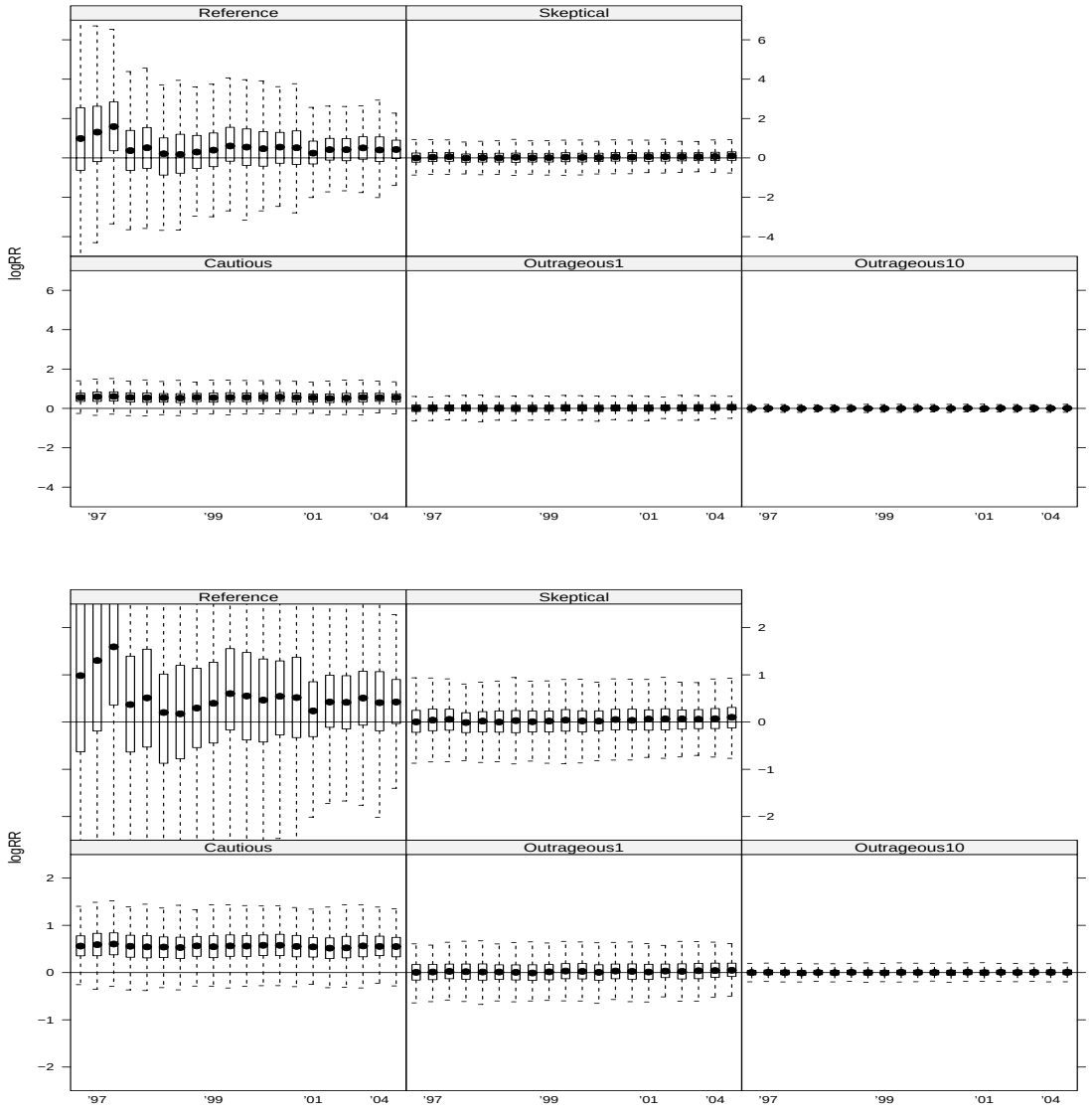


Figure 4.6: Box Plots for the log RR under Poisson data models.

Now focusing on the block effect model, we draw the box plots for the posterior draws of RR's in logarithm scale at various time points and prior distributions in Figure 4.6. The box plot shows that the (log) RR's are above zero with $reference$ prior and $cautious1$ prior so that Vioxx is more like to incur CVT than placebo. Comparing to the similar box plot in Figure 4.3 under Gaussian data model, the RR's under the two prior distributions do not have the upward trend quite as obvious.

Lastly we use the MCMC draws from the block effect model and estimate the probability of the (log)RR as in Section 4.3.2.

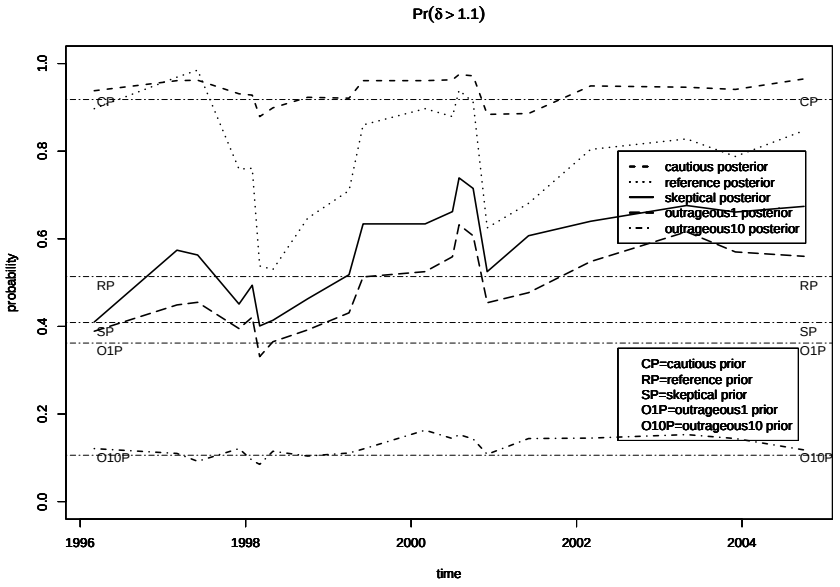


Figure 4.7: Sequential prior and posterior $\Pr(\delta > \delta_L)$ for all Vioxx placebo-controlled trials under Poisson block effect model.

Figure 4.7 shows the prior and posterior probability that $\delta > \delta_L$ under all priors using the Poisson block effect model. Compared to its counterpart using the Gaussian fixed effect model in Figure 4.4, the posterior probabilities remain relatively flat over the time. Consistent with Figure 4.6, the posterior probability that the relative risk was greater than δ_L exceeded 50% under all priors except the "outrageous10" prior after the middle of year 2000.

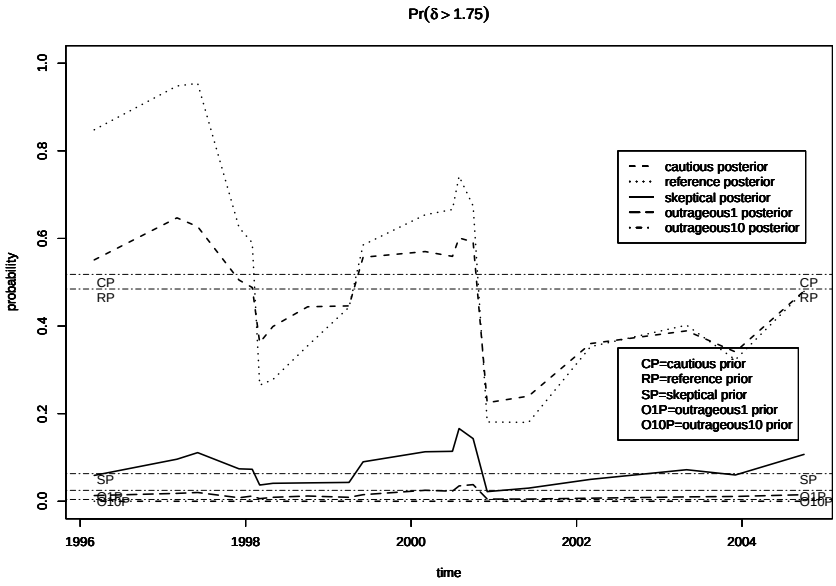


Figure 4.8: Sequential prior and posterior $\Pr(\delta > \delta_U)$ for all Vioxx placebo-controlled trials under Poisson block effect model.

Figure 4.8 shows the posterior probability that $\delta > \delta_U$ over time under Poisson block effect model. The model provides much stronger support for a value of δ as large as 1.75 for cautious prior and reference prior than the Gaussian fixed effect model.

4.4 Discussions and Conclusion

The Poisson data model approach selects a different model from the Gaussian data model. It appears that the models with block configuration are superior. To shed light on this, we separately estimate the Poisson rates under Vioxx arm and placebo arm and plot the confidence interval for the rates in Figure 4.9. From the plot, we observe that under both treatment arms and for all priors, the (log) Poisson rates of ALZ-block differ from the ones of OA-block and APP-block. This shows that block configuration is a complexity that is credible and necessary under Poisson data model since we model data separately on the two

treatment arms. On the other hand, the Gaussian data model only considers the RR's which are similar among the blocks. As a result, we do not need block configuration in the model.

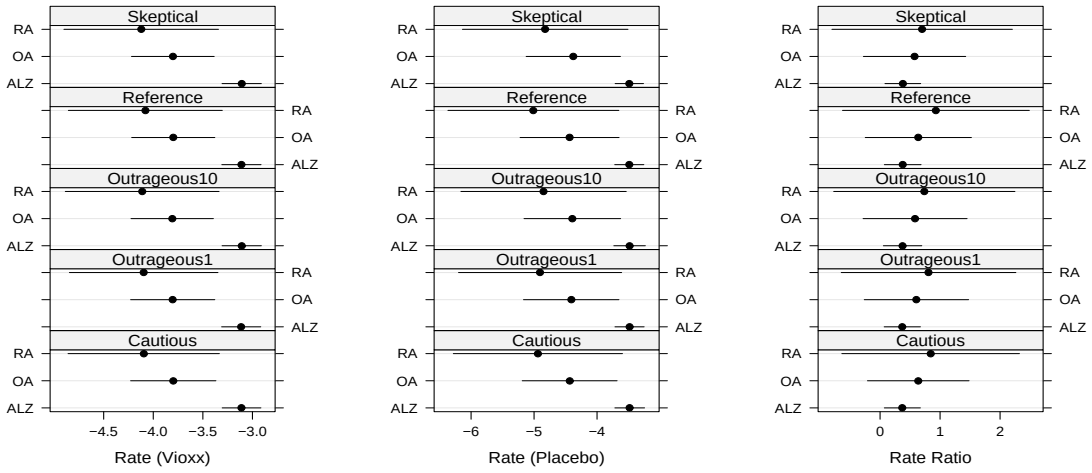


Figure 4.9: Poisson rates under the Vioxx and placebo arm, and log relative risk under Poisson block effect model.

In this chapter, we propose a general and flexible framework for Bayesian sequential meta-analysis to aggregate multiple clinical trials. In practice, there might exist other sources of information, such as subjects' medical records, insurance claim information, prescriptions, etc. This framework can extend to accommodate these kinds of data. With that, we expect more accurate result. On the other hand, with huge amount of data, the sequential meta-analysis is ideally carried out by updating the result with newly available data. We can adopt the approach of particle filter (Ridgeway *et al.*, 2002) to improve the computational efficiency.

Chapter 5
A Latent Model to Detect Overlapping Clusters

5.1 Introduction

The question of whether a spatial pattern is randomly distributed over space after adjusting for a known inhomogeneity has been of interest in many applications such as spatial cluster monitoring of a disease (Molinari *et al.*, 2001; Demattei *et al.*, 2007), surveillance of biological terrorism (Wallenstein and Naus, 2004), detection of unusal pattern clusters in DNA sequencing (Leung *et al.*, 2005), etc.

A traditional statistical method to detect a cluster of events in spatial data is via *Scan Statistics* (Glaz and Balakrishnan, 1999; Glaz *et al.*, 2001; Balakrishnan and Koutras, 2001; Fu and Lou, 2003). The most commonly used scan statistic is the maximum number of cases in a fixed size moving window that scans through the study area. The test based on this scan statistic has been shown to be a generalized likelihood ratio test for a uniform null against a false alternative. A related scan statistic is the diameter of the smallest window that contains a fixed number of cases. Other scan statistics and related likelihood based tests for localized temporal or spatial clustering have been developed, often using a range of fixed window sizes or a range of fixed number of cases (Kulldorff and Nagarwalla, 1995; Naus and Wallenstein, 2004; Dembo and Karlin, 1992; Su *el al.*, 2001).

In addition to classical statistical methods, Bayesian approaches that compute posterior probabilities of potential clusters have also been proposed for cluster detection problems. "Bayesian spatial scan statistic" is proposed for spatial cluster

detection by Neill *et al.* (2006). This method uses a conjugate Gamma-Poisson model instead of Poisson model for the model assumption. Compared with the standard frequentist methods, this Bayesian method can not only incorporate prior information about the size and shape of a cluster, but also the impact of the cluster on the monitored data stream. However, just as with the standard frequentist scan statistics methods, the potential spatial clusters are limited to a finite set of specific choices.

Gangnon and Clayton (2001) developed a "weighted average likelihood ratio" (WALR) statistic with a Bayesian interpretation. This statistic approximates posterior probabilities of a cell being part of the cluster, and in turn helps locate the cluster. Later they (Gangon and Clayton, 2004) further developed two other scan type statistics: a "weighted average likelihood ratio scan" (WALRS) statistic and a "penalized scan statistic".

Some Bayesian methods are based on the "disease mapping framework"(Denison, 2001; Gangon and Clayton, 2000; Gangon and Clayton, 2003; Lawson, 1995) Among them, the Bayesian approach proposed in Gangnon and Clayton (2000) incorporates ideas from image analysis, Bayesian model averaging and model selection. Starting with a saturated model with certain number of clusters, a randomized model search algorithm similar to backwards elimination is proposed by repeatedly merging adjacent components to produce models with high posterior densities.

The above Bayesian methods can both allow for multiple cluster detection and produce estimates for disease rates. However, the potential spatial clusters are limited to the cell divisions. In addition, disease rates are estimated conditional on the estimated clusters. Such conditional estimation may not accurately reflect the uncertainty about the composition of the cluster. Meanwhile, the choice of priors is always challenging.

Although scan statistics procedures have been successful in detecting a single

cluster as well as multiple clusters of fixed size, problems arise for detecting multiple clusters of varying sizes. In recent years, there have been several attempts to overcome the difficulty. A well known approach is a stepwise regression model together with model selection procedures to locate and determine the number of unusually high clustering regions (Demattei *et al.*, 2006; Demattei *et al.*, 2007). These approaches rely on a weighted least square formulation, although the response variable (gaps between incidents) is typically non-Gaussian. Recently, Xie, Sun and Naus (2009) developed a latent cluster model for temporal data which allows the use of the standard likelihood inference for detecting multiple clusters. Sun (2008) extended the temporal cluster detection to spatial data and developed a spatial cluster detection method to simultaneously detect multiple clusters of varying sizes, as well as a significant single cluster. These approaches are based on likelihood inference and they are more efficient in detecting clusters of varying sizes than the weighted least squares approaches. Cheng and Xie (2009) used the spatial latent modeling method (Sun, 2008) to design a robust surveillance system based on a network of mobile sensors for detecting nuclear materials in a metropolitan area.

The latent model approach (Sun, 2008) made an assumption that all the clusters are non-overlapping. This assumption simplifies the theoretical formation as well as its implementation. This chapter extends the approach to the case of overlapping clusters. The rest of the chapter is organized as follows. Section 5.2 details the latent model allowing overlapping clusters. Section 5.3 applies the methodology to nuclear detection in a metropolitan area using mobile sensor network. Section 4.4 concludes the chapter with some discussions.

5.2 Models and Methodology

We first assume that there are k clusters denoted by $I_1, ..., I_k$ in a given two dimensional region, $I = (0, X) \times (0, Y)$. The clusters are assumed to be circles with centers at $\mathbf{O} = (\mathbf{o_1}, \mathbf{o_2}, ..., \mathbf{o_k})$ where $\mathbf{o}_j = (x_j, y_j)^T$ represents the coordinates for the j-th cluster, radii $\mathbf{r} = (r_1, r_2, ..., r_k)^T$ where r_j is the radius for the j-th cluster. The locations $\mathbf{O}$ and sizes of the clusters $\mathbf{r}$ are treated as latent random variables. We assume that centers $\mathbf{O}$ follow a density function $\psi_{\mathbf{o}}(\mathbf{x}) = \psi_{\mathbf{o}}(\mathbf{x}; \lambda_o)$ and radii $\mathbf{r}$ a density function $\psi_r(\mathbf{x}) = \psi_r(\mathbf{x}; \lambda_r)$ with a positive support. One simple example for $\psi_{\mathbf{o}}(\mathbf{x})$ to use is a uniform distribution on I. The common choice of $\psi_r(\mathbf{x})$ is among a truncated exponential, an inverse Gamma or a log-normal distributions. Here λ_o and λ_r are unknown parameters jointed denoted by $\lambda = (\lambda_o, \lambda_r)$.

For convenience, we introduce a random variable δ such that $\{\delta = k\}$ is the event that k clusters occur in the region I. Since the probability $P_\lambda(\delta = k)$ is not explicit, we can use Monte Carlo method to numerically approximate its value. Given $\{\delta = k\}$, the joint conditional likelihood function of $(\mathbf{O}, \mathbf{r})$ is

$$f_\lambda(\mathbf{O}, \mathbf{r}|\delta = k) = \frac{\prod_{j=1}^{k} \{\psi_o(\mathbf{o_j})\psi_r(r_j)\}\mathbf{1}_{\{\delta=k\}}}{P_\lambda(\delta = k)} \tag{5.1}$$

5.2.1 Likelihood Function of Event Points

We assume now that at most two clusters overlap and the observations $\mathbf{y}$=($\mathbf{y_1}$, $\mathbf{y_2}$,..., $\mathbf{y_n}$) are independent identically distributed (iid) samples from a piecewise

uniform density function $f_\theta(\xi|\mathbf{O}, \mathbf{r}, k)$ defined as:

$$f_\theta(\xi|\mathbf{O}, \mathbf{r}, k) = \begin{cases} \alpha_1/c, & \text{if } \xi \in I_1 \text{ and } \eta \notin \bigcup_{j' \neq 1} I_{j'} \\ \quad \cdots\cdots & \\ \alpha_k/c, & \text{if } \xi \in I_k \text{ and } \eta \notin \bigcup_{j' \neq k} I_{j'} \\ g_2(\alpha_1, \alpha_2)/c, & \text{if } \xi \in I_1 \cap I_2; \\ \quad \cdots\cdots & \\ g_2(\alpha_{k-1}, \alpha_k)/c, & \text{if } \xi \in I_{k-1} \cap I_k; \\ 1/c, & \text{if } \xi \notin \bigcup_{j=1}^{k} I_j \end{cases} \tag{5.2}$$

where $\theta = (\alpha^T, \lambda^T)'$ is the collection of all parameters, including parameters $\alpha = (\alpha_1, \ldots, \alpha_k)'$ which are densities inside clusters relative to non-cluster area (or background), and parameters $\lambda = (\lambda_O, \lambda_r)'$ that are associated with random vectors $\mathbf{o}_j$'s and variables r_j's. $g_2(.)$ is the function for aggregating the densities from the two clusters. Different application scenario determines the function g_2. When event counts are additive from the two clusters, we can have $g_2(\alpha_i, \alpha_j) = \alpha_i + \alpha_j$. In the case where events occur when a certain threshold is exceeded from the two clusters, we can have $g_2(\alpha_i, \alpha_j) = max(\alpha_i, \alpha_j)$ for all $i \neq j$. In this study, we assume g_2 is unknown and treat $g_2(\alpha_i, \alpha_j)$ as an additional parameter for the (i, j) pair in θ to be estimated.

Using function $A(.)$ for the area of a region, we can express the normalizing constant c as:

$$c = A(\cap_{j=1}^{k} I_j^c) + \sum_{j=1}^{k} \alpha_j A(I_j \cap (\bigcup_{j' \neq j} I_{j'}^c)) + \sum_{j \neq j'} g_2(\alpha_j, \alpha_{j'}) A(I_j \cap I_{j'}) \tag{5.3}$$

In the special case of non-overlapping clusters, the third term in (5.3) disappears.

The conditional joint density function of the observations $\mathbf{y} = (\mathbf{y_1}, \mathbf{y_2}, ..., \mathbf{y_n})'$ given $\mathbf{O}, \mathbf{r}$ and k is the product of the stepwise uniform density functions in (5.2)

for the n event points:

$$f_\theta(\mathbf{y}|\mathbf{O},\mathbf{r},k) = exp\left\{\sum_{j=1}^{k} Z_j log\alpha_j + \sum_{j\neq j'} Z_{j,j'} log[g_2(\alpha_j,\alpha_{j'})] - nlog(c)\right\} \quad (5.4)$$

where Z_j is the number of points in $\mathbf{y}$ that lie *only* in the region I_j, $Z_{i,j}$ is the number of points in $\mathbf{y}$ that lie in the *intersection* of I_j and $I_{j'}$:

$$Z_j = Z_j(\mathbf{y},\mathbf{O},\mathbf{r},k) = \sum_{i=1}^{n} \mathbf{1}_{\{\mathbf{y_i}\in I_j \cap (\bigcup_{j'\neq j} I^c_{j'})\}} \quad (5.5)$$

$$Z_{j,j'} = Z_{j,j'}(\mathbf{y},\mathbf{O},\mathbf{r},k) = \sum_{i=1}^{n} \mathbf{1}_{\{\mathbf{y_i}\in I_j \cap I_{j'}\}} \quad (5.6)$$

We treat $(\mathbf{y},\mathbf{O},\mathbf{r})$ as *complete* responses and $\mathbf{y}$ as *observed* response. The joint likelihood of the complete data is

$$\begin{aligned} f_\theta(\mathbf{y},\mathbf{O},\mathbf{r},k) &= f_\theta(\mathbf{y}|\mathbf{O},\mathbf{r},k) f_\theta(\mathbf{O},\mathbf{r}|k) P_\lambda(\delta=k) \\ &= \frac{\prod_{j\neq j'}[g_2(\alpha_j,\alpha_{j'})]^{Z_{j,j'}(\mathbf{y},\mathbf{O},\mathbf{r})}}{[c(\mathbf{y},\mathbf{O},\mathbf{r},\theta)]^n} \prod_{j=1}^{k} \{\alpha_j^{Z_j(\mathbf{y},\mathbf{O},\mathbf{r})} \varphi_{\lambda_o}(\mathbf{o_j}) \varphi_{\lambda_r}(r_j)\} \mathbf{1}_{\{\delta=k\}} \end{aligned}$$

5.2.2 Monte-Carlo EM Algorithm for Model Estimation

When the observations are available, we want to determine whether there are any significant clusters and also detect the locations of the potential clusters. For the first problem, we can test a hypothesis H_0: $\alpha_1 = \alpha_2 = ... = 1$ versus H_1: at least one $\alpha_j \neq 1$. The estimation of the model parameters θ solves the second problem. For both tasks, we need to calculate and maximize the *observed* likelihood from the standard likelihood inference procedures:

$$f_\theta(\mathbf{y},k) = \int ... \int f_\theta(\mathbf{y},\mathbf{O},\mathbf{r},\delta=k) d\mathbf{O} d\mathbf{r} \quad (5.7)$$

The integration is difficult to compute directly. To solve the problem, we use an Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977) where we treat $(\mathbf{y},\delta=k)$ as the observed variables:

Step 0. Select a set of starting parameter values $\theta^{(0)}$.

Step 1. (E-step). For $s = 0, 1, 2, \ldots$, calculate the conditional expectation of the complete log-likelihood function, given the observed and $\theta = \theta^{(s)}$: $Q(\theta|\theta^{(s)}) = Q_1(\alpha|\theta^{(s)}) + Q_2(\lambda|\theta^{(s)})$ where

$$\begin{aligned} Q_1(\alpha|\theta^{(\mathbf{s})}) &= \sum_{j=1}^{k} E(Z_j|\mathbf{y}, k, \theta^{(\mathbf{s})}) \log \alpha_j + \sum_{j \neq j'} E(Z_{j,j'}|\mathbf{y}, k, \theta^{(\mathbf{s})}) \log g_2(\alpha_j, \alpha_{j'}) \\ &\quad - nE[\log(c)|\mathbf{y}, k, \theta^{(\mathbf{s})}] \\ Q_2(\lambda|\theta^{(\mathbf{s})}) &= \sum_{j=1}^{k} E\{\log \psi_o(o_j)|\mathbf{y}, k, \theta^{(\mathbf{s})}\} + \sum_{j=1}^{k} E\{\log \psi_r(r_j)|\mathbf{y}, k, \theta^{(\mathbf{s})}\} \end{aligned}$$

Step 2 (M-step). For $s = 0, 1, 2, \ldots$, update the parameter estimates: $\theta^{(s+1)} = (\alpha^{(s+1)}, \lambda^{(s+1)})'$, by maximizing the $Q_1(\alpha|\theta^{(\mathbf{s})})$ and $Q_2(\lambda|\theta^{(\mathbf{s})})$ functions: $\alpha^{(s+1)} = \operatorname{argmax} Q_1(\alpha|\theta^{(\mathbf{s})})$ and $\lambda^{(s+1)} = \operatorname{argmax} Q_2(\lambda|\theta^{(\mathbf{s})})$.

Step 3. Repeat Steps 2 and 3 until $\|\theta^{(s+1)} - \theta^{(s)}\|$ is very small.

In the E-step of the EM algorithm, we need to calculate the conditional expectations in terms of the conditional likelihood of the unobserved data given the observed: $f_\theta(\mathbf{O}, \mathbf{r}|\mathbf{y}, k)$. Since $P_\lambda(\delta = k)$ does not have an explicit form, we approximate the expectations by the Monte Carlo simulation, which relies on the random draws of the multiple clusters based on the observed events points. We turn to Gibbs sampling scheme to generate clusters one at a time based on the fully conditional likelihood of $(\mathbf{o_j}, r_j)$:

$$\begin{aligned} f_\theta((\mathbf{o_j}, r_j) \;|\; (\mathbf{o_l}, r_l), l = 1, 2.., k, l \neq j, \mathbf{y}, k) &\propto f_\theta(\mathbf{O}, \mathbf{r}, \mathbf{y}, k) \\ &\propto \frac{\prod_{j \neq j'} [g_2(\alpha_j, \alpha_{j'})]^{Z_{j,j'}(\mathbf{y}, \mathbf{O}, \mathbf{r})}}{[c(\mathbf{y}, \mathbf{O}, \mathbf{r}, \theta)]^n} \alpha_j^{Z_j(\mathbf{y}, \mathbf{O}, \mathbf{r})} \varphi_{\lambda_o}(\mathbf{o_j}) \varphi_{\lambda_r}(r_j) \mathbf{1}_{\{\delta = k\}} \quad (5.8) \end{aligned}$$

which is impossible for direct simulations. Nevertheless, it is straightforward to simulate a $(\mathbf{o_j}, r_j)$ from $\varphi_{\lambda_o}(\mathbf{o_j})\varphi_{\lambda_r}(r_j)$. Therefore we use an importance sampling method to get around the problem with the weights computed from (5.8).

Suppose $\mathbf{O}^* = (\mathbf{o}_1^*, \ldots, \mathbf{o}_k^*)$ and $\mathbf{r}^* = (r_1^*, \ldots, r_k^*)'$ are a set of Gibbs samples from $f(\mathbf{O}, \mathbf{r}|\mathbf{y}, k, \theta^{(s)})$. They are generated by cycling through simulations from

the fully conditional distributions of A_j given the rest of $\mathbf{o}$'s and r's many times until the Gibbs sampling chain is "burn-in". Repeat the process a large number of times to get M sets of Gibbs samples. The five conditional expectations in the E-step of the EM-algorithm can be evaluated by $\frac{1}{M}\sum_* Z_j^*$, $\frac{1}{M}\sum_* Z_{j,j'}^*$, $\frac{1}{M}\sum_* \log c^*$, $\frac{1}{M}\sum_* \log\{\psi(\mathbf{o}_j^*)\}$, and $\frac{1}{M}\sum_* \log\{\psi(r_j^*)\}$, respectively, where, $\sum_*$ is the summation over M sets of Gibbs samples $\mathbf{O}^*$ and $\mathbf{r}^*$. Z_j^*, $Z_{j,j'}^*$ and c^* are computed with $\mathbf{o}_j$ and r_j values replaced by their corresponding Gibbs sample values $\mathbf{o}_j^*$ and r_j^* in each of the Gibbs sample sets. We list the pseudo code of the EM algorithm in Appendix F.

5.2.3 Likelihood Inference for Tests Related to α's

To test the significance of the estimated parameters, we use either the Wald tests or a likelihood ratio test. Since the Wald test needs the estimation of the variance-covariance matrix, it can be computationally unstable. Therefore we will use the likelihood inference for the two sided tests related to α's. According to likelihood inference, the twice of the log likelihood ratio test statistic is

$$\begin{aligned} R &= 2\log\{\frac{\max_{H_1\cup H_0} f_\theta(\mathbf{y},k)}{\max_{H_0} f_\theta(\mathbf{y},k)}\} = 2\log\{\max_{H_1\cup H_0} f_\theta(\mathbf{y},k)\} - 2\log(\frac{1}{E})^n \\ &= 2\log\{f_{\hat\theta}(\mathbf{y},k)\} + 2\log(E)^n \\ &= 2\log\int\int f_{\hat\theta}(\mathbf{y}|\mathbf{O},\mathbf{r},k) f_{\hat\theta}(\mathbf{O},\mathbf{r}|k) d\mathbf{O} d\mathbf{r} + 2\log P_{\hat\lambda}(\delta = k) + 2n\log(E) \end{aligned}$$

where $\hat\theta = (\hat\alpha, \hat\lambda)^T$ are MLEs estimated from the aforementioned EM algorithm and E is the area of the study region I.

Suppose for a moment that we know how to simulate $\mathbf{O}^{**} = (\mathbf{o}_1^{**}, \ldots, \mathbf{o}_{k+1}^{**})$ and $\mathbf{r}^{**} = (r_1^{**}, \ldots, r_k^{**})$ from $f(\mathbf{O},\mathbf{r}|k)$ when $\theta = \hat\theta$, and we have M sets of such simulated $\mathbf{O}^{**}$ and $\mathbf{r}^{**}$ samples. By Monte-Carlo approximation, the test statistic R can be approximated by

$$R^{**} = 2\Bigg[\log\{\frac{1}{M}\sum_{**} f(\mathbf{y}|\mathbf{O}^{**},\mathbf{r}^{**},k)\} + \log P_{\hat\lambda}(\delta = k) + n\log(E)\Bigg],$$

where $\sum_{**}$ is the summation over the M sets of $\mathbf{O}^{**}$ and $\mathbf{r}^{**}$ samples. Based on the likelihood inference we know that R is asymptotically χ^2 distributed with k^* degrees of freedom, where k^* is the number of parameters to be estimated. So, compare R^{**} with the χ^2_k distribution we can perform a formal test for $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_{k^*} = 1$ versus H_1 : at least one $\alpha_j \neq 1$.

We again use the Gibbs sampling approach to simulate $\mathbf{O}^{**} = (\mathbf{o}_1^{**}, \ldots, \mathbf{o}_k^{**})$ and $\mathbf{r}^{**} = (r_1^{**}, \ldots, r_k^{**})$ from $f(\mathbf{O}, \mathbf{r}|k)$ from any set of given parameter values θ. Note that,

$$f((\mathbf{o}_j, r_j)|(\mathbf{o}_l, r_l), l = 1, 2, \ldots, k, l \neq j, \mathbf{y}, k) \propto \psi_{\mathbf{o}}(\mathbf{o}_j)\psi_r(r_j)\mathbf{1}_{(\delta=k)}$$

Similar to the steps in Section 5.2.2, we can use importance sampling method to simulate from the truncated distributions.

5.2.4 Identification of Cluster Regions

If a cluster is significant (i.e. $\alpha_j \neq 1$), we often want to determine the cluster region. Note that the jth cluster A_j is determined by the center $\mathbf{o}_j$ and the radius r_j. Their conditional expectations given $\mathbf{y}$ and k (posterior means in the context of Bayesian paradigm) are $\mathrm{E}\{\mathrm{o}_j|\mathbf{y}, k\}|_{\theta=\hat{\theta}}$ and $\mathrm{E}\{\mathrm{r}_j|\mathbf{y}, k\}|_{\theta=\hat{\theta}}$. The cluster center $\mathbf{o}_j$ and the radius r_j can be simply estimated by $\frac{1}{M}\sum_* \mathbf{o}_j^*$ and $\frac{1}{M}\sum_* r_j^*$ respectively. Here $\sum_*$ is the summation over the M sets of Gibbs samples in the last iteration of the EM algorithm.

An alternative approach is to use the medians of the M sets of o_j^* and r_j^* to estimate o_j and r_j, respectively. Since the distribution may not be symmetric, this median method may provide more accurate estimators.

5.2.5 Determination of the Unknown Number of Clusters

In previous sections, we assume that the number of clusters is known. It is rarely true in reality. We now describe a model selection approach to determine the

number of clusters from the observed data. We propose to use both AIC and BIC criteria. The AIC criterion (Akaike, 1974) is a commonly used model selection method developed based on the Kullback-Leibler information between the candidate models and the true model. Schwarz (1978) obtains the BIC procedure by using Bayes estimators and a fixed penalty for choosing the wrong dimension. Both criteria minimize an expression that consists of a term that measures model fit plus a term that penalizes model complexity. In our context, a direct application of the AIC and BIC rules yields

$$\begin{aligned}\mathrm{AIC}(k) &= -2\log f_\theta(\mathbf{y},k)+2k\\ &= -2\log[\int\int f_\theta(\mathbf{y}|\mathbf{O},\mathbf{r},k)f_\theta(\mathbf{O},\mathbf{r}|k)d\mathbf{O}d\mathbf{r}]-2\log P_\lambda(\delta=k)+2k\end{aligned}$$

and

$$\begin{aligned}\mathrm{BIC}(k) &= -2\log f_\theta(\mathbf{y},k)+k\log(n)\\ &= -2\log[\int\int f_\theta(\mathbf{y}|\mathbf{O},\mathbf{r},k)f_\theta(\mathbf{O},\mathbf{r}|k)d\mathbf{O}d\mathbf{r}]-2\log P_\lambda(\delta=k)+k\log(n)\end{aligned}$$

Often $n > e^2 = 7.389$, the BIC method places more penalty against a large number of clusters than the AIC method.

The parameters θ are unknown. To compute the criteria, these parameters should be replaced by their estimators $\hat{\theta} = \hat{\theta}(k)$ that are obtained by the EM/MCMC algorithm proposed in the previous section. Furthermore, the formula involves integrations that do not have explicit forms. We numerically evaluate their values. From the previous section, we know how to simulate $\mathbf{O}^{**} = (\mathbf{o}_1^{**},\ldots,\mathbf{o}_k^{**})$ and $\mathbf{r}^{**} = (r_1^{**},\ldots,r_k^{**})$ from $f(\mathbf{O},\mathbf{r}|k)$ when $\theta = \hat{\theta}$. By Monte-Carlo approximation, the criterion AIC(k) can be approximated by

$$\widehat{\mathrm{AIC}}(k) = -2\log\big[\frac{1}{M}\sum_{**} f(\mathbf{y}|\mathbf{O}^{**},\mathbf{r}^{**},k)\big]-2\log P_{\hat{\lambda}}(\delta=k)+2k, \qquad (5.9)$$

and BIC(k) criterion can be approximated by

$$\widehat{\mathrm{BIC}}(k) = -2\log\big\{\frac{1}{M}\sum_{**} f(\mathbf{y}|\mathbf{O}^{**},\mathbf{r}^{**},k)\big\}-2\log P_{\hat{\lambda}}(\delta=k)+k\log(n), \quad (5.10)$$

where $\sum_{**}$ are the summations over M sets of repeatedly simulated o^{**}'s and r^{**}'s. The k to be chosen is in fact the one with the smallest corresponding $\widehat{\text{AIC}}(k)$ or $\widehat{\text{BIC}}(k)$ value.

Denote $\mathcal{K}$ as a pre-selected set of k's. We want this set small for computing purpose but large enough to cover all potential choices of the correct number of clusters. For each fixed k in $\mathcal{K}$, we apply the EM/MCMC algorithm to obtain the parameters estimates. We then use either AIC or BIC rule to determine the optimal number of clusters. For the chose k, we can determine the cluster regions.

5.2.6 Incorporate Background Information

Sometimes, the underlying sensors density or the radioactive background for the study region is not uniform. We need to incorporate this information into the clustering procedure. Let the function $B()$ denote the background density information. The piecewise density function (5.2) is modified as

$$f_\theta(\xi|\mathbf{O},\mathbf{r},k) = \begin{cases} \frac{\alpha_1 B(\xi)}{c^*}, & \text{if } \xi \in I_1 \text{ and } \xi \notin \bigcup_{j'\neq 1} I_{j'}; \\ \quad \cdots\cdots & \\ \frac{\alpha_k B(\xi)}{c^*}, & \text{if } \xi \in I_k \text{ and } \xi \notin \bigcup_{j'\neq k} I_{j'}; \\ \frac{g_2(\alpha_1,\alpha_2) B(\xi)}{c^*}, & \text{if } \xi \in I_1 \cap I_2; \\ \quad \cdots\cdots & \\ \frac{g_2(\alpha_{k-1},\alpha_k) B(\xi)}{c^*}, & \text{if } \xi \in I_{k-1} \cap I_k; \\ \frac{B(\xi)}{c^*}, & \text{if } \xi \notin \bigcup_{j=1}^k I_j \end{cases} \tag{5.11}$$

where c^* is the modified normalizing constant

$$c^* = \int\int_{\cap_{j=1}^k I_j^c} B(\xi)d\xi + \sum_{j=1}^k \alpha_j \int\int_{I_j \cap (\bigcup_{j'\neq j} I_{j'}^c)} B(\xi)d\xi + \sum_{j\neq j'} g_2(\alpha_j,\alpha_{j'}) \int\int_{I_j \cap I_{j'}} B(\xi)d\xi$$

This model reduces to Equation (5.2) when $B(\xi) \propto 1$, or uniform background. The steps for cluster detection using this model are similar to these with (5.2).

5.3 Application to Nuclear Detection

Threats to national security have become more dynamic and complex in the past decade due to global terrorism, increased opposition to U.S. interests, greater pursuit of nuclear power and expanded access by adversaries to sophisticated technologies and materials. Among all the threats, nuclear attacks are arguably the most devastating. They can cause severe losses and casualties in human lives as well as long term and large scale damage to infrastructure. As the result, there have been growing concerns regarding the prospect of transporting, storing and detonating nuclear materials or dirty bombs in the populous metropolitan areas. Thus it becomes increasingly vital to have sophisticated nuclear detection systems deployed in major cities. Proactive monitoring and detection via pervasive surveillance is crucial to detect and thwart the malicious attacks (Carpenter *et al.*, 2010). To help achieve this goal, we propose a mobile sensor network and use the latent model approach to detect nuclear sources in a metropolitan area.

5.3.1 Mobile Sensor Network

A surveillance network of mobile sensors has the prototype designs as follows:

- Nuclear sensors and Global Position System (GPS) tracking devices are installed on a large number of vehicles such as taxicabs, police vehicles, fire trucks, and buses.

- The sensors and GPS devices constantly send detection and location information to a central command center. These signals are marked onto a map of a metropolitan area under surveillance.

- Real time analysis is performed at the command center using sophisticated statistical algorithms including the latent modeling method to detect and pinpoint nuclear sources.

In such a network, when vehicles with sensors move within a certain range of a nuclear source, the radiation energy from the source will trigger the sensor devices to send out wireless signals to a central command center along with the positions of the sensors. The random movement and extensive coverage of the vehicles provide a constant surveillance of nuclear materials. With large quantities, the mobile sensors do not need to be of high accuracy, since the failure of a small portion of them will not significantly affect the effectiveness of the surveillance coverage due to sensors' random movements. Moreover, it is almost impossible to tamper with such a network of devices.

Due to many attractive characteristics of sensor networks, there have been many studies and applications of the sensor networks in military and civil applications including surveillance, smart homes, remote environment monitoring. See Akyildize *et al.* (2002a, b) for a recent survey. Much of the research devotes to sensor placement, sensor reorganization and communications. In the area of radiation detection, the idea of using massive mobile sensors has been adopted and tested by the Radiation Laboratory at Purdue University (Purdue, 2008). They use a network of cell phones with GPS capabilities to detect and track radiation. The noise and false positive detection problems are tackled by setting and tuning the solid state devices. A multi-sensor nuclear threat detection problem was studied in Hochbaum (2008) using a combinatorial network flow algorithm.

Since the sensor signals are not 100% accurate, there are always false alarms or missed detections. For example, a sensor might display positive readings when there is no such signal, or fail to detect a real signal nearby. In this study we consider probabilistic models for sensor reading and source detection. These models are generalized to include multiple sources with different aggregation rules.

5.3.2 Models for Nuclear Intensity, Sensor Reading and Detection

We consider a nuclear source in this chapter as a portable nuclear device transported by an individual via trucks or bags (FEMA, 2008). As the nuclear radiation starts from a source, the total energy stays as a constant due to the Conservation Law of Energy. For simplicity, we assume that radiation travels in spherical waves. Let $z(r)$ denote the intensity at distance r. The total energy remaining a constant for all r is $4\pi r^2 z(r)$, where $4\pi r^2$ is the surface area of the sphere with radius r. As the radius increases by a factor of k, the surface area of the sphere will increase by a factor of k^2. As a result, the radiation intensity z decreases by the inverse square of the distance r: $z(r) = c/r^2$, where the constant c is a factor related to the total energy of the source (Wein, 2006). Since the nuclear detection device is triggered by radiation intensity, getting closer to the nuclear source will better the chance for detection. The ubiquitous nature of the mobile sensor network takes advantage of this property.

Let S denote the status of the sensor's reading with the value of 1 for a positive reading and 0 otherwise. We describe S with a threshold model:

$$S = \mathbf{1}_{\{z(r)\geq d\}} = \mathbf{1}_{\{c/r^2\geq d\}} \tag{5.12}$$

where d is a threshold for detection and $\mathbf{1}_{\{.\}}$ is the indicator function. That is, if the intensity $z(r)$ at the sensor location is greater than the threshold d, the sensor will detect the source; otherwise the sensor reports a negative reading.

In practice there might be multiple nuclear sources, whose energy levels and positions will jointly determine the reading status of a sensor. In this chapter, we use a generic approach without assumptions about the nature of the different sources and how the sensors are activated. In some situations we can have simple forms for the detection model. Let Ω be the number of sources, c_ω be energy factor of the ωth source, r_ω be the distance from the sensor to this source. When each

source has different energy spectrum, they can activate a sensor independently as long as the energy threshold is exceeded. Now the threshold model is

$$S = \mathbf{1}_{\{max_{\omega\in\{1,..\Omega\}} z_\omega \geq d\}} = \mathbf{1}_{\{max_{\omega\in\{1,..\Omega\}} c_\omega/r_\omega^2 \geq d\}} \tag{5.13}$$

In the other case where the energies from the sources are within same spectrum of frequencies, the aggregation of intensities from all sources at the senor location is: $z_{total} = \sum_{\omega=1}^{\Omega} c_\omega/r_\omega^2$. From the threshold model (5.12), the reading S can be determined by:

$$S = \mathbf{1}_{\{z_{total}\geq d\}} = \mathbf{1}_{\{\sum_{\omega=1}^{\Omega} c_\omega/r_\omega^2 \geq d\}} \tag{5.14}$$

As with any detection device, a nuclear sensor may not be 100% accurate. The sensor errors can be from the variability in the manufacturing process, routine wear and tear, missing scheduled maintenance and calibrations, and undetected malfunctions. In addition, random traces of weak environmental nuclear signals can also trigger false alerts. For example, a person who just went through a radioactive therapy or a bag of cat litter can set off alarms. We regard such sources as *trivial* ones as they are weak and last a short period of time. Furthermore the wireless signals from the mobile sensor to the control center may incur transmission errors. We use the two parameters *sensitivity* and *specificity* to assess the average performance of a sensor device. In the context of nuclear detection, sensitivity, denoted as η, presents the probability of detecting nuclear sources where there are indeed such materials. Specificity, denoted as ζ, is the probability of not detecting any nuclear materials where there in fact do not exist any. Let D be the binary indicator of a sensor detecting a true nuclear source, D equal to 1 for the positive detection and 0 otherwise. We have $\eta = P(D = 1|S = 1)$ and $\zeta = P(D = 0|S = 0)$.

The quality control characteristics of a sensor, false negative rate (FNR) and false positive rate (FPR), can be expressed in η and ζ as: $FNR = P(D = 0|S = 1) = 1 - \eta$ and $FPR = P(D = 1|S = 0) = 1 - \zeta$. Then the probability of

detecting a nuclear source is:

$$\begin{aligned} P(D=1) &= P(D=1|S=1)P(S=1) + P(D=1|S=0)P(S=0) \\ &= (1-\zeta) + (\zeta+\eta-1)P(S=1). \end{aligned} \tag{5.15}$$

Under the perfect scenario where both η and ζ are 1, the detection D is the same as the reading S.

The above models are related to the cluster detection problem in statistics. The threshold model (5.12) can be expressed as $S = \mathbf{1}_{\{A\}}$, where $A = \{r \leq (c/d)^{1/2}\}$ is a sphere, or a circle on a 2-dimensional map, centered at the nuclear source and with radius $R = (c/d)^{1/2}$. The ratio of the probabilities of a positive reading inside and outside the set A is $P(D=1|A)/P(D=1|\bar{A}) = P(D=1|S=1)/P(D=1|S=0) = \eta/(1-\zeta)$. In the case when both FNR and FPR are less than 25%, for instance, we have the ratio greater than 3. That is, the sensor is 3 times more likely to report a positive signal ($D=1$) inside A than inside $\bar{A}$ with moderate accuracy. This type of statement matches the definition of a spatial cluster in the statistical literature, in which the clusters are defined as areas within which an incident of interest is more likely to happen (i.e., with a higher probability of happening per squared unit) than outside these areas. In our setting, an incident of interest is an alert signal with $D=1$.

5.3.3 Simulation of a Mobile Sensor Network

A street grid based simulation tool was developed in Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) at Rutgers University to study the traffic patterns in a metropolitan area. The tool supports multiple turnable parameters such as the numbers of streets in either horizontal or vertical direction, the size of street blocks, several types of vehicles and their numbers, etc. A snapshot of the simulation tool is in Figure (5.1).

We use the simulation tool in this chapter to obtain the random positions of

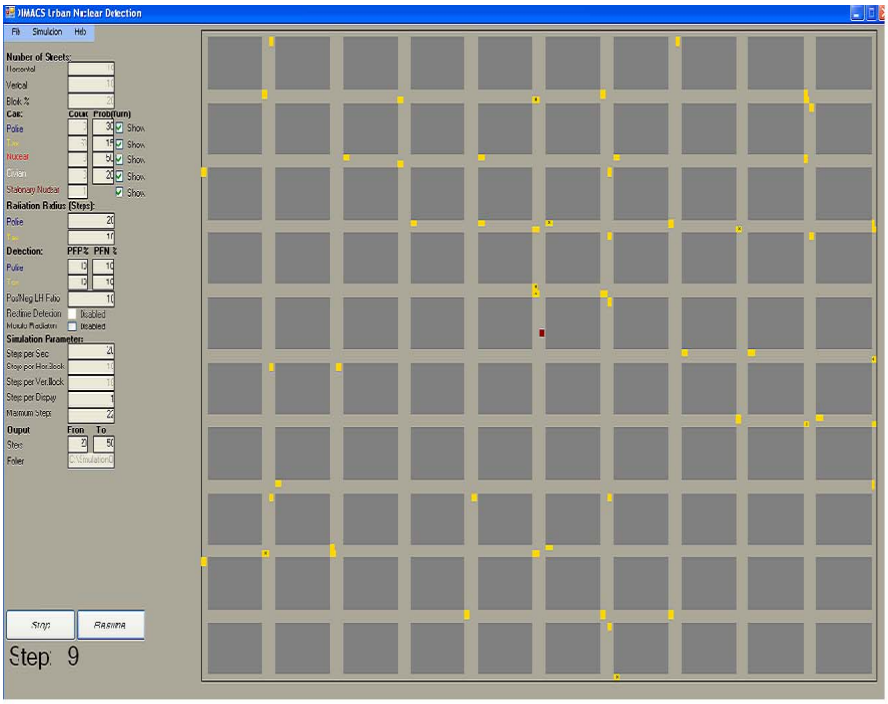


Figure 5.1: Snapshot of the simulation tool

vehicles installed with nuclear detectors. The models in Section 5.3.2 determine a probability of detection for each sensor. With it, we can simulate the positions of the sensors which might indicate an existence of true nuclear source. Because of the random errors, we apply the latent modeling approach to detect the sources with statistical significance. We declare a correct detection if the clusters are statistically significant and cover the true locations of the nuclear sources.

5.3.4 Power of Detecting Overlapping Clusters

The usefulness of a mobile sensor network is to effectively detect a nuclear source. Detection power is a measurement of such ability under a set of network design parameters such as the number, qualities and detection range of sensors. Using simulations, the power is defined as the ratio of detecting the true clusters over

the total number of simulations. In this section we conduct simulation studies to demonstrate that the proposed network and method can effectively detect multiple and overlapping nuclear sources.

We aim to design a mobile sensor network in an area with similar size of Downtown Manhattan. The study region is set henceforth to an area of 25×25 blocks. Each block is a 20 by 20 square with one unit representing 10 feet in real distance. Since total number of taxi cabs in the New York City is about 13,000, it is reasonable to assume that there are 1,500 to 2,000 cabs operating during the rush hours in the busy downtown area.

The set of the network parameters is selected as: sensor detection range of 200 feet (the same size of a block), error rates of (5%, 5%)(false positive rate and false negative rate), number of taxi cabs of 1,500. We set two sources with various distances resulting in overlapping clusters, close-by but non-overlapping clusters, and far-away clusters. In each case, the two sources jointly activate sensors from the threshold model (5.14) where we set $c_1 = c_2 = 20$, $d = 1$. We repeat 500 simulations and compute how many times (in percentage) the proposed algorithm can correctly detect a randomly placed nuclear source as an estimate of the statical power of detection $P(D = 1|S = 1)$.

We use set $\mathbf{o_1}$ and $\mathbf{o_2}$ for the centers, r_1 and r_2 for the radii for the two sources. Let $\mathbf{x_i}$ and r_i be the center and radius for the i-th cluster (i=1,2). The detection is correct if the two clusters are significant from LRT test and the sources are inside the clusters, i.e., $d_i = ||\mathbf{x_i} - \mathbf{o_i}|| \leq r_i$ for i=1 and 2. When the LRT test result is significant and one of the two sources is insides the clusters, we declare that one single source is detected.

With each set of event points, we use the overlapping, non-overlapping with preset $K = 2$, and non-overlapping with preset $K = 1$ algorithms to detect the true clusters. For each set of two true clusters, we calculated powers for covering both of the sources and either one of the two sources. We then use four empirical

measures to assess the accuracy of the estimated clusters' locations: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Sensitivity is the proportion of the event points inside the true clusters, that are inside the estimated clusters. Specificity is the proportion of the event points outside the true clusters, that are outside the estimated clusters. PPV is the proportion of the event points inside the estimated clusters, that are inside the true clusters. NPV is the proportion of the event points outside the estimated clusters, that are outside the true clusters. The closer these measures are to one, the more accurate the estimated clusters are.

Table 5.1: Powers and empirical measures to assess the accuracy of the clusters' locations.

Case	d/(r1+r2)		Overlapping preset k=2	Non-overlapping preset k=1	Non-overlapping preset k=2
1	0.2				
		Power (both)	98.8 %	98.0 %	7.8 %
		Power (only 1)	0.6%	0.0 %	82.8 %
		Sensitivity	0.982 (0.051)	0.993 (0.040)	0.742 (0.211)
		Specificity	0.942 (0.028)	0.937 (0.032)	0.965 (0.026)
		PPV	0.666 (0.138)	0.653 (0.141)	0.726 (0.175)
		NPV	0.998 (0.005)	0.999 (0.003)	0.972 (0.021)
2	0.7				
		Power (both)	99.2 %	98.6 %	36.6 %
		Power (only 1)	0.1 %	1.0 %	60.8%
		Sensitivity	0.977 (0.070)	0.995 (0.049)	0.849 (0.166)
		Specificity	0.905 (0.034)	0.886 (0.037)	0.929 (0.035)
		PPV	0.535 (0.130)	0.494 (0.120)	0.580 (0.156)
		NPV	0.997 (0.007)	1.000 (0.003)	0.983 (0.019)
3	1.4				
		Power (both)	86.8 %	82.8 %	79.2 %
		Power (only 1)	13.0 %	17.0 %	12.6 %
		Sensitivity	0.896 (0.132)	0.916 (0.111)	0.854 (0.182)
		Specificity	0.936 (0.022)	0.917 (0.027)	0.941 (0.024)
		PPV	0.674 (0.087	0.656 (0.077)	0.714 (0.089))
		NPV	0.961 (0.024)	0.987 (0.016)	0.978 (0.027)
4	2.5				
		Power (both)	32.2 %	0.4 %	82.8 %
		Power (only 1)	65.4 %	96.0 %	13.4 %
		Sensitivity	0.624 (0.154)	0.553 (0.120)	0.905 (0.151)
		Specificity	0.980 (0.028)	0.996 (0.010)	0.995 (0.008)
		PPV	0.888 (0.120)	0.972 (0.057)	0.977 (0.039)
		NPV	0.924 (0.032)	0.912 (0.027)	0.984 (0.022)

Table 5.1 summarizes the results of cluster detection and testing. When we have overlapping clusters (Case 1 and 2), the overlapping algorithm achieves better detection power than the non-overlapping one. The non-overlapping method

can only detect the two true clusters as one (with preset $K = 1$). When the distance between the two clusters increases to a slightly non-overlapping case (Case 3), the non-overlapping with $K = 2$ increases its detection power dramatically, while the overlapping method still has its advantage over the non-overlap with $K = 1$ method. In the totally non-overlapping case (Case 4), overlapping method is not doing as good as the non-overlapping $K = 2$ one which reflects the true clusters' configuration. The four empirical measures of the detection do not differ much among the three algorithms in all cases.

In summary, this set of simulations show that overlapping algorithm perform well when the true clusters indeed overlap with each other. When the true clusters are far apart, non-overlapping code is the right method to use and can achieve better detection power.

5.4 Discussions and Conclusions

We extend and generalize the modeling and inference framework of the latent model approach (Sun, 2008) to accommodate cases of overlapping clusters. The method is applied to a mobile sensor network with consistent and pervasive surveillance for nuclear materials in major cities. Simulation studies suggest that our method can achieve better detection powers than the original algorithm when the true clusters overlap.

Even though we have made an assumption that at most two clusters overlap, we can extend our method to the case of three or more clusters straightforward, at least in theory. We expect the extension will be more computationally expensive. Therefore it might not make practically sense to pursue this. Sun (2008) dealt with spatial clusters with shape of ellipses. This is certainly more general than the circle clusters that we have in this chapter. Our future research will use the ellipse in the overlapping cluster detection.

Chapter 6
Conclusions

We have discussed some non-standard missing data problems in clinical trials and cluster detection in this thesis.

First, we used a model-base approach to handle non-ignorable drop-out in longitudinal clinical trials. The sensitivity study was conducted to assess the robustness of the inference to different assumptions of missing data mechanisms. We provided insight to the dependence via a simulation study.

Second, we proposed a flexible Bayesian framework to manage non-compliance coupled with non-ignorable drop-out in clinical trials. Adopting the Rubin Causal Model, we calculated the complier average causal effect to better estimate the treatment effect and reconciled different results from various models through Bayesian model averaging perspective. In the future, we can further enhance the efficiency of MCMCM methodology and Bayes factor algorithm. With that, we can extend the framework to a longitudinal setting.

Third, we developed a Bayesian sequential meta-analysis framework. In relation to the first research topics to provide better and more accurate assessment of a new treatment, this flexible framework aggregates results from all available studies and updates the findings as new data becomes available. In practice, there exist sources of information other than clinical trials, such as medical records, insurance claim and prescription information, we can extend this framework to accommodate these different types of data. To handle the huge amount of data computationally, we can adopt particle filter approach to sequentially accumulate

results.

Last, we extended and generalized a latent model approach to overlapping clusters. We applied this method to a mobile sensor network with consistent and pervasive surveillance for nuclear materials in major cities. In the future, we can further extend the approach to shapes more general than circles (such as ellipses) for the clusters.

Appendix A
Gibbs Simpler for CACE

We illustrate in details the MCMC steps to calculate CACE under Model 1 in Chapter 3 (3.2.2). Under Bayesian context, the parameter set $\mathbb{P}$ contains both model parameters $\Phi = \{\alpha^1, \beta^1, \sigma^1, \gamma^1, \alpha^0, \beta^0, \sigma^0, \gamma^0\}$ and missing data $D^{miss} = \{D^1_{miss}, D^0_{miss}, Y^1_{miss}, Y^0_{miss}, M^1_{miss}, M^0_{miss}\}$. We use vectors for these variables. For example, $\alpha^v = \{\alpha^v_1, \alpha^v_2\}$, $D^1_{miss} = \{D^1_{miss,i} : i = 1, 2, ...N_{D^1_{miss}}\}$ where $N_{D^1_{miss}}$ is the number of missing data in D^1. Observed data D^{obs} is $\{D^v_{obs}, D^p_{obs}, Y^v_{obs}, Y^p_{obs}, M^v_{obs}, M^p_{obs}\}$.

We use random-walk Metropolis-Hasting algorithm to generate samples from target distribution $\pi(x)$. Given the current state $x^{(t)}$,

- Generate $y = x^{(t)} + \epsilon$ and ϵ is a random pertubation, for example, $\epsilon \sim U[-0.1, 0.1]$ when y is a continuous varaible.

- Simulate $u \sim uniform[0, 1]$ and set

$$x^{(t+1)} = \begin{cases} y & \text{if } u \leq \frac{\pi(y)}{\pi(x^{(t)})} \\ x^{(t)} & \text{otherwise} \end{cases}$$

To get a Bayesian estimate of the $CACE$ from (3.2), we need samples from the posterior distribution of the missing variables in D^v, D^p, Y^v and Y^p through $MCMC$. Gibbs sampling achieves this with the fully conditional posterior distribution of a parameter η given the rest of the paramters (denoted as $\mathbb{P}_{-\eta}$) and data:

$$f(\eta|\mathbb{P}_{-\eta}, \mathbb{D}^{obs}) = Cf(\eta, \mathbb{P}_{-\eta}, \mathbb{D}^{obs}) = Cf(\mathbb{P}, \mathbb{D}^{obs}) \tag{A.1}$$

where the normalizing constant C is independent of η because

$$C = \frac{1}{\int f(\mathbb{P}, \mathbb{D}^{obs}) d\eta} \tag{A.2}$$

Therefore C will cancel out in the Gibbs sampling procedure when we take the ratio of the $f(\eta^{new}|\mathbb{P}_{-\eta}, \mathbb{D}^{obs})$ to $f(\eta^{current}|\mathbb{P}_{-\eta}, \mathbb{D}^{obs})$. As the result, we can use the joint pdf of data and parameters as the target distribution $\pi(x)$ instead of the conditional distributions which depends on the parameter of interest. We further write the joint pdf as:

$$\begin{aligned} f(\mathbb{P}, \mathbb{D}^{obs}) &= f(\mathbb{D}^{obs}, \mathbb{D}^{miss}, \Phi) = \prod f(\mathbb{D}_i^{obs}, \mathbb{D}_i^{miss}|\Phi) f(\Phi) \\ &= \prod f(D_i^1, D_i^0, Y_i^1, Y_i^0, M_i^1, M_i^0|\Phi) f(\Phi) \end{aligned} \tag{A.3}$$

The Bayesian models for $(D_i^1, D_i^0, Y_i^1, Y_i^0, M_i^1, M_i^0)$ determine the further factorization. Under the Bayesian Network 1:

$$\begin{aligned} f(\mathbb{D}_i^{obs}, \mathbb{D}_i^{miss}|\Phi) = &\ f(D_i^1|\Phi) f(Y_i^1|D_i^1, \Phi) f(M_i^1|D_i^1, Y_i^1, \Phi) \\ &\times f(D_i^0|D_i^1, \Phi) f(Y_i^0|D_i^0, Y_i^1, \Phi) f(M_i^0|D_i^0, Y_i^0, M_i^1, \Phi) \end{aligned}$$

The prior distribution of Φ is

$$f(\Phi) = f_N(\alpha^{\mathbf{1}}) f_N(\alpha^{\mathbf{0}}) f_N(\beta^{\mathbf{1}}) f_N(\beta^{\mathbf{0}}) f_N(\gamma^{\mathbf{1}}) f_N(\gamma^{\mathbf{0}}) f_G(\sigma^1) f_G(\sigma^0)$$

where each of the parameters from $\{\alpha^{\mathbf{1}}, \alpha^{\mathbf{0}}, \beta^{\mathbf{1}}, \beta^{\mathbf{0}}, \gamma^{\mathbf{1}}, \gamma^{\mathbf{0}}\}$ takes a noninformative prior of $N(0, 10^4)$ denoted by $f_N(.)$, and σ^v, σ^p take Gamma(0.01, 100) denoted by $f_G(.)$.

Initialize all the parameters in Φ and $\mathbb{D}^{miss}$. Iterate t=1,2.... update or retain each element in Φ in the order of α_1^v, α_2^v, β_1^v, β_2^v, β_3^v, γ_1^v, γ_2^v, γ_3^v, α_1^p, α_2^p, α_3^p, β_1^p, β_2^p, β_3^p, β_4^p, γ_1^p, γ_2^p, γ_3^p, γ_4^p, σ^v, and σ^p. Denote the parameter to be μ. Denote $\Phi^{[t-1]}$ to be the parameter set taking values the previous iteration, and $\Phi_{-\mu}^{[t-1]}$ to be subset of $\Phi^{[t-1]}$ less the element μ.

Then update or retain each element in D^{miss} in the order of D_{obs}^v, Y_{obs}^v, M_{obs}^v, M_{obs}^p, D_{obs}^p, and Y_{obs}^p. Denote the variable to be W_i for i-th subject. To simplify

notations, we denote the collection of $(D_i^v, Y_i^v, M_i^v, D_i^p, Y_i^p, M_i^p)$ for the subject to be D_i, and $D_i^{[t-1]}$ to take the values from the previous iteration, and $D_{i,-W_i}^{[t-1]}$ to be subset less the element of W_i. The Gibbs sampling steps are:

1. Generate $\mu^{[t]} \sim U[\mu^{[t-1]} - 0.1, \mu^{[t-1]} + 0.1]$ or $W_j^{[t]} \sim U[W_j^{[t-1]} - 0.1, W_j^{[t-1]} + 0.1]$. (The range of the uniform distribution is to be adjusted for better sampling efficiency.)

2. Calculate ratio of the joint pdf in (A.3) given the current $\mu^{[t]}$ or $W_j^{[t]}$ plus the rest of the original parameter/data set to the pdf given the previous parameter/data set:

   - For $\mu^{[t]}$, the ratio r is:

   $$r = \frac{\prod_i f(\mathbb{D}_i|\mu^{[t]}, \Phi_{-\mu}^{[t-1]}) f(\mu^{[t]}, \Phi_{-\mu}^{[t-1]})}{\prod_i f(\mathbb{D}_i|\Phi^{[t-1]}) f(\Phi^{[t-1]})} \tag{A.4}$$

   Take α_1^v for example:

   $$r = \frac{\prod_i f(D_i^v|(\alpha_1^v)^{[t]}, (\alpha_2^v)^{[t-1]}) f_N((\alpha_1^v)^{[t]})}{\prod_i f(D_i^v|(\alpha_1^v)^{[t-1]}, (\alpha_2^v)^{[t-1]}) f_N((\alpha_1^v)^{[t-1]})} \tag{A.5}$$

   - For $W_j^{[t-1]}$,the pdfs of the subjects other than j and the prior pdf for the parameters cancel out in the ratio. The ratio r is:

   $$r = \frac{f(W_j^{[t]}, \mathbb{D}_{j,-W_j}^{[t-1]}|\Phi^{[t]})}{f(\mathbb{D}_j^{[t-1]}|\Phi^{[t]})} \tag{A.6}$$

3. Generate $u \sim U[0, 1]$. If $u \geq r$, reject the value of $\mu^{[t]}$ or $W_j^{[t]}$ and keep the previous values as the current ones.

4. Now calculate $CACE_t$ for t^{th} iteration from (3.2).

After discarding burn-in of first 1000 iteration, we can calculate the posterior mean of $CACE$ by averaging the remaining $CACE_t$ with t=1001, 1002.....

Appendix B
Methods for Computing Bayes Factors

Discrete Random Variable Case

Heckerman (1996) summerizes methods of computing exact marginal likelihood for discrete random variables. From multinomial sampling, the observed variable X is discrete, having r possible states $\{x^1, ..., x^r\}$. The likelihood function is

$$p(X = x^k|\theta, M) = \theta_k, \qquad k = 1, ..., r$$

where $\theta = \{\theta_1, ..., \theta_r\}$ are the parameters subject to $\sum_{k=1}^{r} \theta_k = 1$, M is for model specification. The sufficient statistics for data $\mathbb{D} = \{N_1, ..., N_r\}$ where N_k is the number of times $X = x_k$ from data. Using the simple conjugate prior of the Dirichlet distribution:

$$p(\theta|M) = Dir(\theta|\alpha_1, ..., \alpha_r) = \frac{\Gamma(\alpha)}{\prod_{k=1}^{r} \Gamma(\alpha_k)} \prod_{k=1}^{r} \theta_k^{\alpha_k - 1}$$

where $\alpha = \sum_{k=1}^{r} \alpha_k$, and $\alpha_k > 0$. Let $N = \sum_{k=1}^{r} N_k$. Then the marginal likelihood $p(\mathbb{D}|M)$ is deduced as

$$p(D|M) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} \prod_{k=1}^{r} \frac{\Gamma(\alpha_k + N_k)}{\Gamma(\alpha_k)} \tag{B.1}$$

Extending the above to n nodes indexed by i $(i = 1, ..n)$, each of which has r_i possible states (or values), q_i parent nodes indexed by j, the marginal likelihood of the data is just the product of the marginal likelihood for each $i - j$ pair from (B.1):

$$p(\mathbb{D}|M) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha_{ijk} + N_{ijk})}{\Gamma(\alpha_{ijk})} \tag{B.2}$$

We can use (B.2) to calculate Bayesian factor between two models which have discrete reponses. We can use value of 1 for all α_k (k=1,...,r) which responds to a non-informative prior distribution.

For continuous response, we can first standardize the variable (substracting mean, then dividing by standard deviation) and approximate the resulting variable into a discrete one (e.g., a binary one using 0 as the dividing value).

Laplace Method for Continuous Response

The marginal likelihood under a model M_i (i=1,2) is:

$$[\mathbb{D}|M_i] = \int [\mathbb{D}, \Phi|M_i] d\Phi = \int [\mathbb{D}|\Phi, M_i][\Phi|M_i] d\Phi \tag{B.3}$$

Usually the Bayes factor is intractable and thus must be computed by numerical methods. Monte Carlo method is one of the choices and it straightforward since we can easily obtain draws from $[\Phi|M_i]$ - the prior distribution of the parameter set. However, in practice this approach is problematic. One reason is that when sample sizes are moderate or large, the integrand becomes highly peaked around its maximum and the mothod is not efficient as we will have difficulty finding the region where the integrand mass accumulates. The second reason is that some problems are of high dimension (i.e. there are 21 parameters in Graphical Model 1 in Section (3.2.2)).

Laplace's method first maximizes the joint distribution $[\mathbb{D}, \Phi|M_i]$ at its mode $\Phi^m_{M_i}$. Use function h to denote the log of the integrand in (B.3) and expand it by a Taylor series about $\Phi^m_{M_i}$.

$$\begin{aligned}
[\mathbb{D}|M_i] &= \int exp[h(\Phi|\mathbb{D}, M_i)] d\Phi \\
&\approx exp[h(\Phi^m_{M_i}|\mathbb{D}, M_i)] \int exp[1/2(\Phi - \Phi^m_{M_i})(-I_{M_i})(\Phi - \Phi^m_{M_i})] d\Phi \\
&= [\mathbb{D}|\Phi^m_{M_i}, M_i][\Phi^m_{M_i}|M_i] \int exp[1/2(\Phi - \Phi^m_{M_i})(-I_{M_i})(\Phi - \Phi^m_{M_i})] d\Phi
\end{aligned} \tag{B.4}$$

Use Hessian matrix I_{M_i} to denote $-h^{''}(\Phi|\mathbb{D}, M_i)|_{\Phi=\Phi^m_{M_i}}$ and substitute (B.4) into (B.3):

$$\begin{aligned}[\mathbb{D}|M_i] &= \int exp[h(\Phi|\mathbb{D}, M_i)]d\Phi \\ &\approx exp[h(\Phi^m_{M_i}|\mathbb{D}, M_i)] \int exp[1/2(\Phi - \Phi^m_{M_i})(-I_{M_i})(\Phi - \Phi^m_{M_i})]d\Phi \\ &= [\mathbb{D}|\Phi^m_{M_i}, M_i][\Phi^m_{M_i}|M_i] \int exp[1/2(\Phi - \Phi^m_{M_i})(-I_{M_i})(\Phi - \Phi^m_{M_i})]d\Phi \end{aligned} \tag{B.5}$$

The integrand in (B.5) is proportional to an multivariate normal with A_{M_i} as a precision matrix. Assume the dimension of Φ is d. This leads to a Laplace approximation of the marginal distribution:

$$[\mathbb{D}|M_i] \approx [\mathbb{D}|\Phi^m_{M_i}, M_i][\Phi^m_{M_i}|M_i](2\pi)^{d/2}|I_{M_i}|^{-0.5} \tag{B.6}$$

As the result, the ratio of the two marginal distributions or the Bayes factor is

$$\frac{[\mathbb{D}|M_2]}{[\mathbb{D}|M_1]} \approx \frac{[\mathbb{D}|\Phi^m_{M_2}, M_2][\Phi^m_{M_2}|M_2]|I_{M_2}|^{-0.5}}{[\mathbb{D}|\Phi^m_{M_1}, M_1][\Phi^m_{M_1}|M_1]|I_{M_1}|^{-0.5}} \tag{B.7}$$

BIC Method for Continuous Response

An alternative is to adopt an approximation to the Bayes factor, the Bayesian Information Criterion(BIC) is equal to $logp(y|\hat{\theta}, M) - \frac{p}{2}log(n)$. The first term is the familiar probability of the data given the model, computed at the value $\hat{\theta}$ that maixmizes this probability. The second term promotes model parsimony by penalizing models with increased model complexity (large p) and sample size.

Therefore the logorithm of the Bayes factor is

$$log\frac{[\mathbb{D}|M_2]}{[\mathbb{D}|M_1]} \approx log[\mathbb{D}|\Phi^m_{M_2}, M_2] - log[\mathbb{D}|\Phi^m_{M_1}, M_1] - log(n)(p_2 - p_1)/2 \tag{B.8}$$

Appendix C
Extension to Longitudinal Study for CACE

We extend Model 1 in Section 3.2.3 to a longitudinal setting in Figure C.1. The subscript of each node denotes the time period. Each node at one time is assumed to depend on its value of the previous period.

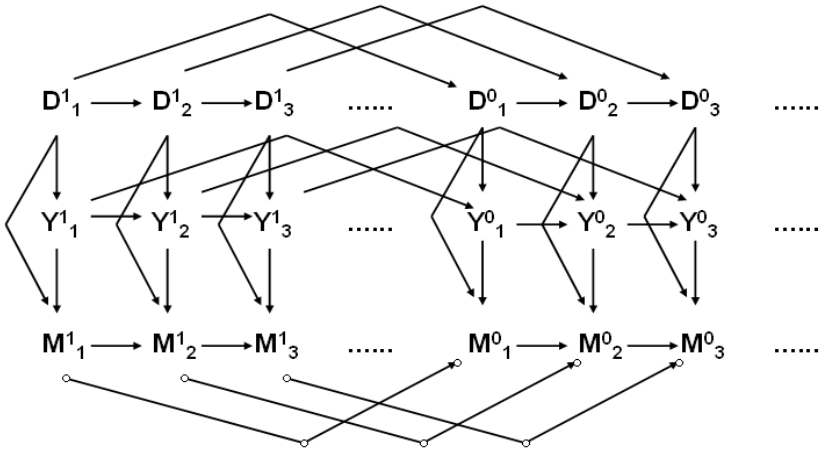


Figure C.1: Bayesian graphical model for a longitudinal trial setting.

We work out the conditional likelihood of each node as follows:

1. $D^v_{i,j}$ - Bayesian Graphic Model: $(D^v_{i,j-1}, X) \longrightarrow D^v_{i,j}$

$$logit(P(D^v_{i,0} = 1)) = \alpha^v_{1,0} + \alpha^v_{2,0} HC_i$$

$$logit(P(D^v_{i,j} = 1)) = \alpha^v_{1,j} + \alpha^v_{2,j} HC_i + \alpha^v_{3,j} D^v_{i,j-1} \quad j \geq 2 \qquad \text{(C.1)}$$

2. $D^p_{i,j}$ - Bayesian Graphic Model: $(D^p_{i,j-1}, X) \longrightarrow D^p_{i,j}$

$$logit(P(D^p_{i,0} = 1)) = \alpha^p_{1,0} + \alpha^p_{2,0} HC_i$$

$$logit(P(D^p_{i,j} = 1)) = \alpha^p_{1,j} + \alpha^p_{2,j} HC_i + \alpha^p_{3,j} D^p_{i,j-1} \quad j \geq 2 \qquad \text{(C.2)}$$

3. $Y^v_{i,j}$ (SBP) - Bayesian Graphic Model: $(D^v_{i,j}, Y^v_{i,j-1}, X) \longrightarrow Y^v_{i,j}$

$$\begin{aligned} Y^v_{i,1} &= \beta^v_{1,1} + \beta^v_{2,1} HC_i + \beta^v_{3,1} + \beta^v_{4,1} D^v_{i,1} + \epsilon^v_{i,1} \\ Y^v_{i,j} &= \beta^v_{1,j} + \beta^v_{2,j} HC_i + \beta^v_{3,j} j + \beta^v_{4,j} D^v_{i,j} + \beta^v_{5,j} Y^v_{i,j-1} + \epsilon^v_{i,j}, \quad j \geq 2 \end{aligned} \tag{C.3}$$

where $\epsilon^v_{i,j} \sim N(0, \sigma^2_v)$.

4. $Y^p_{i,j}$ (SBP) - Bayesian Graphic Model: $(D^p_{i,j}, Y^p_{i,j-1}, X) \longrightarrow Y^p_{i,j}$

$$\begin{aligned} Y^p_{i,1} &= \beta^p_{1,1} + \beta^p_{2,1} HC_i + \beta^p_{3,1} + \beta^p_{4,1} D^p_{i,1} + \epsilon^p_{i,1} \\ Y^p_{i,j} &= \beta^p_{1,j} + \beta^p_{2,j} HC_i + \beta^p_{3,j} j + \beta^p_{4,j} D^p_{i,j} + \beta^p_{5,j} Y^p_{i,j-1} + \epsilon^p_{i,j}, \quad j \geq 2 \end{aligned} \tag{C.4}$$

where $\epsilon^p_{i,j} \sim N(0, \sigma^2_p)$.

5. $M^v_{i,j}$ - Bayesian Graphic Model: $(D^v_{i,j}, Y^v_{i,j-1}, Y^v_{i,j}, M^v_{i,j-1}, X) \longrightarrow M^v_{i,j}$

$$\begin{aligned} logit(P(M^v_{i,1} = 1)) &= \gamma^v_{1,1} + \gamma^v_{2,1} HC_i + \gamma^v_{3,1} D^v_{i,1} + \delta^v Y^v_{i,1} \\ logit(P(M^v_{i,j} = 1)) &= \gamma^v_{1,j} + \gamma^v_{2,j} HC_i + \gamma^v_{3,j} D^v_{i,j} + \gamma^v_{4,j} M^v_{i,j-1} \\ &\quad + \delta^v_{prev} Y^v_{i,j-1} + \delta^v Y^v_{i,j} \qquad j \geq 2 \end{aligned} \tag{C.5}$$

6. $M^p_{i,j}$ - Bayesian Graphic Model: $(D^p_{i,j}, Y^p_{i,j-1}, Y^p_{i,j}, M^p_{i,j-1}, X) \longrightarrow M^p_{i,j}$

$$\begin{aligned} logit(P(M^p_{i,1} = 1)) &= \gamma^p_{1,1} + \gamma^p_{2,1} HC_i + \gamma^p_{3,1} D^p_{i,1} + \delta^p Y^p_{i,1} \\ logit(P(M^p_{i,j} = 1)) &= \gamma^p_{1,j} + \gamma^p_{2,j} HC_i + \gamma^p_{3,j} D^p_{i,j} + \gamma^p_{4,j} M^p_{i,j-1} \\ &\quad + \delta^p_{prev} Y^p_{i,j-1} + \delta^p Y^p_{i,j} \qquad j \geq 2 \end{aligned} \tag{C.6}$$

The population Complier Average Causal Effect $(CACE)$ at period j is

$$CACE_j = E_i(Y^v_{i,j} - Y^p_{i,j} | D^v_{i,s} = 1, D^p_{i,s} = 0, 1 \leq s \leq J) \tag{C.7}$$

which is estimated by

$$\widehat{CACE}_j = \frac{1}{\sum_{i=1}^N \left[\prod_{s=1}^J \widehat{D^v_{i,s}}(1 - \widehat{D^p_{i,s}})\right]} \sum_{i=1}^N \left[(\widehat{Y^v_{i,j}} - \widehat{Y^p_{i,j}}) \prod_{s=1}^J \widehat{D^v_{i,s}}(1 - \widehat{D^p_{i,s}})\right] \tag{C.8}$$

Appendix D

Posterior Distribution for δ in Random Effect Model under Gaussian Data Model

In Section 4.2.3, we have random effect model using Gaussian data model:

$$\begin{aligned} r_i|\delta_i &\propto N(\delta_i, \sigma_i^2), \quad \sigma_i^2 = 4/m_i \\ \delta_i|\delta, \tau^2 &\propto N(\delta, \tau^2) \end{aligned} \tag{D.1}$$

For illustration, assume there are three time periods ($i = 1, 2, 3$). At the end of first time period, we observe r_1. The joint posterior distribution of the parameters $(\delta, \delta_1, \delta_2, \delta_3, \tau^2)$ given r_1 is:

$$\begin{aligned} [\delta, \delta_1, \delta_2, \delta_3, \tau^2|r_1] &\propto [r_1, \delta, \delta_1, \delta_2, \delta_3, \tau^2] \\ &\propto [r_1|\delta, \delta_1, \delta_2, \delta_3, \tau^2][\delta, \delta_1, \delta_2, \delta_3, \tau^2] \\ &\propto [r_1|\delta_1][\delta_1|\delta, \tau^2][\delta_2|\delta, \tau^2][\delta_3|\delta, \tau^2][\delta, \tau^2] \end{aligned} \tag{D.2}$$

Integrating δ_2 and δ_3 from the above, we obtain the joint posterior distribution of $(\delta, \delta_1, \tau^2)$ given r_1 as:

$$[\delta, \delta_1, \tau^2|r_1] \propto [r_1|\delta_1][\delta_1|\delta, \tau^2][\delta, \tau^2] \tag{D.3}$$

Joint posterior distribution for the hyper parameters δ and τ^2

Let $\hat{\delta_1} = (\delta/\tau^2 + \delta_1/\sigma_1^2)/(1/\tau^2 + \sigma_1^2)$ and $1/V_1^2 = 1/\tau^2 + 1/\sigma_1^2$, we can integrate δ_1 from (D.3) directly to obtain:

$$\begin{aligned}
[\delta, \tau^2|r_1] &\propto [\delta, \tau^2] \int [r_1|\delta_1][\delta_1|\delta, \tau^2] d\delta_1 \\
&\propto [\delta, \tau^2] \int e^{-\frac{1}{2\sigma_1^2}(r_1-\delta_1)^2} \frac{1}{\tau} e^{-\frac{1}{2\tau^2}(\delta-\delta_1)^2)} d\delta_1 \\
&\propto [\delta, \tau^2] \frac{1}{\sqrt{\sigma_1^2+\tau^2}} e^{-\frac{1}{2(\sigma_1^2+\tau^2)}(\delta-r_1)^2} \int \frac{1}{V_1} e^{-\frac{1}{2V_1^2}(\delta_1-\hat{\delta_1})^2} d\delta_1 \\
&\propto [\delta|\tau^2][\tau^2] N(\delta|r_1, \sigma_1^2+\tau^2) \qquad \text{(D.4)}
\end{aligned}$$

This is similar to (5.18) of Gelman at el. (2004).

Posterior distribution of δ given τ^2

Holding τ^2 constant and using the family of priors for $\delta|\tau^2 \propto N(\delta_0, \sigma_0^2)$, we obtain the conditional posterior distribution of δ given τ^2 and r_1 as:

$$\begin{aligned}
[\delta|\tau^2, r_1] &\propto N(\delta|\delta_0, \sigma_0^2) N(\delta|r_1, \sigma_1^2+\tau^2) \\
&\propto N(\delta|\hat{\delta} = \frac{r_1/(\sigma_1^2+\tau^2) + \delta_0/\sigma_0^2}{1/(\sigma_1^2+\tau^2) + 1/\sigma_0^2}, \sigma_\delta^2 = \frac{1}{1/(\sigma_1^2+\tau^2) + 1/\sigma_0^2}) \qquad \text{(D.5)}
\end{aligned}$$

Posterior distribution of τ^2

Similar to (5.21) of Gelman *et al.* (2004), we use (D.4) and (D.5) to get $[\tau^2|r_1]$ analytically:

$$\begin{aligned}
[\tau^2|r_1] &= \frac{[\delta, \tau^2|r_1]}{[\delta|\tau^2, r_1]} \\
&\propto \frac{[\tau^2] N(\delta|\delta_0, \sigma_0^2) N(\delta|r_1, \sigma_1^2+\tau^2)}{N(\delta|\hat{\delta}, \sigma_\delta^2)} \qquad \text{(D.6)} \\
&\propto [\tau^2] \frac{\sigma_\delta}{\sqrt{\sigma_1^2+\tau^2}} e^{-\frac{1}{2\sigma_0^2}(\hat{\delta}-\delta_0)^2 - \frac{1}{2(\sigma_1^2+\tau^2)}(\hat{\delta}-r_1)^2} \qquad \text{(D.7)}
\end{aligned}$$

Since $[\tau^2|r_1]$ does not depend on the value of δ, we obtain (7) by substituting $\delta = \hat{\delta}$ in (6).

Computation of Posterior Probability $P(\delta > c|r_1)$ for some constant c

The steps are 1) drawing posterior samples from $[\delta, \tau^2|r_1]$; 2) retaining only the component of δ; 3) counting the proportion that the value is grater than c.

Starting with a flat prior for τ,

- Use (7) to draw samples from $[\tau^2|r_1]$ (rejection method or numerical inverse CDF method).
- For each sample of τ, draw δ according to (5).
- repeat the above two steps to get a large number of samples of δ.

Appendix E
Some Oversized Tables

			Vioxx		Placebo		
Study Block	Study No.	LPO	Events	PYR	Events	PYR	Log RR (SE)
RA	017	5/21/1997	1	8	0	7	-0.2878 (1.64)
	068	9/10/1998	1	49	0	24	0.8718 (1.89)
	097	6/6/2000	0	137	0	62	-0.0847 (2.13)
	096	7/21/2000	4	97	0	58	combined with 103
	098	7/6/2000	0	11	1	12	combined with 103
	103	7/6/2000	0	44	0	45	0.8756 (0.96)
	All RA		6	345	1	208	
OA	010	2/8/1996	2	16	0	7	1.265 (1.85)
	029	2/5/1997	3	46	0	16	1.5308 (1.64)
	033	11/18/1997	1	66	1	9	-1.4887 (1.19)
	040	1/1/1998	2	72	0	11	1.2159 (2.81)
	044	2/18/1998	3	154	0	52	combined with 45
	045	2/18/1998	2	157	3	61	-0.4802 (0.69)
	058	4/1/1998	1	21	0	6	0.7881 (2.22)
	083	2/9/2000	0	21	0	21	-0.040 (2.00)
	085	3/3/1999	1	61	0	28	0.843 (1.91)
	090	5/17/1999	5	56	0	27	2.115 (1.78)
	112	9/8/2000	0	104	0	15	-0.121 (2.87)
	116	6/22/2000	1	54	0	15	0.712 (2.17)
	136	2/5/2002	1	95	1	201	0.522 (1.16)
	219	11/28/2003	0	18	1	8	combined with 220
	220	11/24/2003	0	18	0	8	-1.519 (1.50)
	All OA		22	959	6	485	
OA/RA	All OA/RA		28	1304	7	692	
ALZ	078	4/23/2003	75	1623	52	1762	0.445 (0.18)
	091	11/30/2000	13	369	14	381	-0.0398 (0.38)
	126	5/30/2001	11	193	7	197	0.448 (0.47)
	All ALZ		99	2185	73	2341	
APPROVe	122	9/8/2004	75	5700	49	5828	0.444 (0.18)

Table E.1: A subset of the placebo-controlled Vioxx trials. "LPO" is the date when the last patient concluded the trial. "PYR" is patient years at risk. Events are investigator reported cardiovascular thrombotic events. "Log RR" is the log relative risk.

	Fixed Effect Model					Block Effect Model				
Study No.	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
RA Block	—	—	—	—	—	0.31(0.24)	0.40(0.23)	0.37(0.25)	0.13(0.31)	0.28(0.24)
OA Block	—	—	—	—	—	0.25(0.22)	0.33(0.21)	0.30(0.23)	0.03(0.24)	0.21(0.22)
ALZ Block	—	—	—	—	—	0.34(0.13)	0.38(0.12)	0.37(0.13)	0.26(0.15)	0.32(0.13)
122	—	—	—	—	—	0.36(0.14)	0.41(0.13)	0.39(0.14)	0.29(0.17)	0.34(0.14)
Overall	0.33(0.10)	0.39(0.10)	0.37(0.11)	0.11(0.01)	0.30(0.10)	0.29(0.15)	0.40(0.14)	0.36(0.18)	0.05(0.07)	0.25(0.15)
	Random Effect Model					Random Block Effect Model				
Study No.	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
017	0.28(0.24)	0.36(0.22)	0.33(0.24)	0.04(0.31)	0.24(0.25)	0.16(0.82)	0.30(0.82)	0.25(0.84)	0.07(0.82)	0.13(0.82)
068	0.29(0.25)	0.37(0.22)	0.34(0.25)	0.06(0.30)	0.25(0.25)	0.32(0.85)	0.47(0.84)	0.41(0.86)	0.22(0.85)	0.29(0.85)
097	0.30(0.24)	0.38(0.22)	0.35(0.24)	0.07(0.31)	0.27(0.24)	0.20(0.92)	0.35(0.92)	0.29(0.94)	0.09(0.92)	0.16(0.92)
096,098,103	0.32(0.23)	0.39(0.21)	0.36(0.23)	0.12(0.30)	0.28(0.23)	0.51(0.68)	0.61(0.67)	0.57(0.69)	0.44(0.68)	0.48(0.68)
RA Blk	—	—	—	—	—	0.25(0.45)	0.43(0.44)	0.36(0.50)	0.13(0.44)	0.21(0.44)
010	0.31(0.25)	0.39(0.23)	0.36(0.25)	0.10(0.31)	0.27(0.25)	0.23(0.68)	0.35(0.69)	0.29(0.69)	0.14(0.67)	0.20(0.67)
029	0.32(0.25)	0.40(0.24)	0.37(0.25)	0.11(0.31)	0.28(0.25)	0.27(0.71)	0.39(0.72)	0.33(0.72)	0.18(0.69)	0.24(0.70)
033	0.26(0.26)	0.34(0.23)	0.31(0.25)	-0.02(0.32)	0.21(0.26)	-0.20(0.64)	-0.12(0.65)	-0.16(0.65)	-0.27(0.63)	-0.22(0.64)
040	0.29(0.23)	0.38(0.22)	0.34(0.24)	0.06(0.31)	0.26(0.24)	0.18(0.77)	0.30(0.78)	0.24(0.79)	0.08(0.75)	0.15(0.76)
044,045	0.26(0.24)	0.34(0.22)	0.30(0.24)	-0.00(0.28)	0.22(0.25)	-0.12(0.48)	-0.05(0.49)	-0.09(0.49)	-0.17(0.47)	-0.14(0.48)
058	0.31(0.25)	0.39(0.23)	0.36(0.25)	0.09(0.31)	0.28(0.25)	0.17(0.69)	0.29(0.70)	0.23(0.70)	0.08(0.68)	0.14(0.68)
083	0.29(0.24)	0.37(0.22)	0.34(0.23)	0.05(0.31)	0.25(0.24)	0.10(0.71)	0.21(0.72)	0.15(0.72)	0.01(0.69)	0.07(0.70)
085	0.31(0.25)	0.39(0.23)	0.36(0.25)	0.09(0.32)	0.27(0.25)	0.18(0.67)	0.29(0.68)	0.24(0.68)	0.09(0.65)	0.15(0.66)
090	0.32(0.25)	0.40(0.24)	0.37(0.26)	0.12(0.33)	0.29(0.26)	0.36(0.71)	0.48(0.72)	0.42(0.73)	0.27(0.70)	0.33(0.70)
112	0.29(0.24)	0.37(0.22)	0.34(0.24)	0.06(0.30)	0.25(0.24)	0.09(0.72)	0.21(0.73)	0.15(0.74)	-0.00(0.70)	0.06(0.71)
116	0.30(0.27)	0.38(0.25)	0.35(0.27)	0.07(0.33)	0.26(0.28)	0.20(0.73)	0.31(0.74)	0.26(0.75)	0.11(0.72)	0.17(0.73)
136	0.31(0.22)	0.38(0.21)	0.35(0.22)	0.09(0.29)	0.27(0.23)	0.21(0.59)	0.31(0.60)	0.26(0.60)	0.13(0.58)	0.18(0.59)
219,220	0.26(0.26)	0.35(0.24)	0.31(0.26)	-0.01(0.32)	0.22(0.27)	-0.13(0.65)	-0.04(0.67)	-0.08(0.67)	-0.21(0.64)	-0.15(0.65)
OA Blk	—	—	—	—	—	0.13(0.35)	0.25(0.35)	0.19(0.37)	0.03(0.33)	0.10(0.34)
078	0.35(0.13)	0.40(0.13)	0.38(0.13)	0.28(0.16)	0.33(0.13)	0.42(0.17)	0.43(0.17)	0.42(0.17)	0.41(0.17)	0.42(0.17)
091	0.25(0.21)	0.32(0.20)	0.29(0.21)	0.04(0.22)	0.22(0.21)	0.11(0.33)	0.14(0.33)	0.12(0.33)	0.08(0.33)	0.10(0.33)
126	0.33(0.19)	0.39(0.18)	0.37(0.19)	0.16(0.24)	0.29(0.20)	0.38(0.35)	0.42(0.35)	0.40(0.35)	0.35(0.35)	0.37(0.35)
ALZ Blk	—	—	—	—	—	0.25(0.31)	0.35(0.30)	0.31(0.32)	0.18(0.31)	0.23(0.31)
122	0.36(0.14)	0.40(0.14)	0.39(0.14)	0.28(0.17)	0.34(0.14)	0.40(0.17)	0.44(0.17)	0.43(0.17)	0.38(0.17)	0.39(0.17)
ALL	0.30(0.14)	0.38(0.13)	0.35(0.15)	0.06(0.07)	0.26(0.13)	0.17(0.24)	0.42(0.23)	0.32(0.38)	0.02(0.07)	0.12(0.19)

Table E.2: Estimated relative risks using Gaussian data model. P1-skeptical prior, P2-cautious prior, P3-reference prior, P4-outrageous1 prior, and P5-outrageous10 prior.

	Fixed Effect Model					Block Effect Model				
Study No.	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
RA Blk	—	—	—	—	—	0.70(0.75)	0.84(0.74)	0.93(0.78)	0.74(0.75)	0.81(0.73)
OA Blk	—	—	—	—	—	0.57(0.43)	0.64(0.42)	0.64(0.44)	0.58(0.43)	0.61(0.43)
ALZ Blk	—	—	—	—	—	0.38(0.15)	0.37(0.15)	0.38(0.15)	0.38(0.16)	0.37(0.15)
122	—	—	—	—	—	0.45(0.18)	0.44(0.18)	0.45(0.18)	0.43(0.18)	0.44(0.18)
ALL	0.33(0.10)	0.39(0.10)	0.37(0.11)	0.11(0.06)	0.30(0.10)	0.10(0.31)	0.54(0.29)	0.50(0.76)	0.01(0.07)	0.06(0.22)
	Random Effect Model					Random Block Effect Model				
Study No.	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
017	0.47(0.87)	0.64(0.84)	0.68(0.90)	0.30(0.83)	0.39(0.81)	1.35(1.66)	1.53(1.70)	1.59(1.68)	1.35(1.64)	1.23(1.70)
068	0.36(0.80)	0.57(0.78)	0.49(0.83)	0.11(0.76)	0.24(0.73)	0.99(1.49)	1.24(1.52)	1.28(1.61)	1.03(1.58)	0.99(1.46)
097	-0.10(0.78)	0.15(0.79)	0.12(0.89)	-0.33(0.75)	-0.13(0.73)	0.18(1.58)	0.35(1.53)	0.33(1.50)	0.08(1.58)	0.14(1.53)
096,098,103	0.53(0.64)	0.66(0.65)	0.66(0.63)	0.41(0.61)	0.45(0.61)	1.13(1.00)	1.20(0.99)	1.26(1.07)	1.18(0.98)	1.13(0.96)
RA Blk	—	—	—	—	—	0.72(1.01)	0.96(0.96)	0.99(1.08)	0.69(0.99)	0.66(0.98)
010	0.78(0.81)	0.95(0.82)	0.92(0.82)	0.57(0.80)	0.70(0.80)	1.65(1.42)	1.68(1.36)	1.76(1.36)	1.73(1.31)	1.71(1.31)
029	0.76(0.72)	0.95(0.78)	0.99(0.78)	0.61(0.75)	0.72(0.71)	1.57(1.20)	1.69(1.26)	1.71(1.21)	1.61(1.28)	1.67(1.33)
033	-0.14(0.76)	0.04(0.77)	-0.02(0.79)	-0.32(0.75)	-0.25(0.76)	-0.24(1.23)	-0.19(1.26)	-0.23(1.23)	-0.21(1.16)	-0.25(1.19)
040	0.36(0.74)	0.56(0.76)	0.56(0.81)	0.19(0.70)	0.33(0.74)	0.94(1.28)	1.03(1.18)	1.04(1.27)	0.96(1.20)	1.01(1.26)
044,045	-0.13(0.54)	-0.02(0.55)	-0.04(0.58)	-0.26(0.52)	-0.17(0.53)	-0.16(0.70)	-0.12(0.73)	-0.11(0.75)	-0.13(0.70)	-0.12(0.69)
058	0.39(0.80)	0.62(0.83)	0.61(0.85)	0.24(0.74)	0.30(0.78)	1.09(1.43)	1.12(1.32)	1.20(1.34)	1.05(1.40)	1.13(1.39)
083	0.17(0.84)	0.48(0.82)	0.41(0.88)	-0.04(0.79)	0.09(0.82)	0.61(1.40)	0.63(1.40)	0.74(1.43)	0.49(1.38)	0.59(1.42)
085	0.27(0.77)	0.52(0.81)	0.48(0.82)	0.13(0.70)	0.20(0.73)	0.76(1.25)	0.94(1.27)	0.96(1.29)	0.85(1.22)	0.89(1.28)
090	1.13(0.76)	1.31(0.76)	1.23(0.79)	0.95(0.71)	1.05(0.70)	2.04(1.17)	2.13(1.15)	2.23(1.15)	2.03(1.11)	2.13(1.17)
112	-0.20(0.82)	0.04(0.82)	-0.02(0.88)	-0.45(0.84)	-0.30(0.81)	-0.07(1.37)	0.04(1.35)	0.06(1.40)	-0.04(1.37)	0.03(1.30)
116	0.26(0.76)	0.43(0.80)	0.46(0.83)	0.04(0.71)	0.19(0.77)	0.72(1.25)	0.85(1.26)	0.78(1.31)	0.74(1.33)	0.72(1.28)
136	0.41(0.68)	0.60(0.69)	0.57(0.72)	0.25(0.68)	0.33(0.68)	0.74(1.02)	0.77(0.97)	0.78(0.99)	0.73(0.97)	0.79(0.98)
219,220	-0.17(0.81)	-0.02(0.79)	-0.05(0.83)	-0.41(0.83)	-0.23(0.82)	-0.44(1.29)	-0.29(1.25)	-0.33(1.35)	-0.43(1.32)	-0.40(1.31)
OA Blk	—	—	—	—	—	0.63(0.63)	0.77(0.64)	0.77(0.66)	0.66(0.64)	0.69(0.61)
078	0.46(0.18)	0.48(0.18)	0.46(0.18)	0.45(0.18)	0.45(0.18)	0.45(0.18)	0.46(0.18)	0.44(0.18)	0.44(0.18)	0.44(0.18)
091	0.03(0.38)	0.09(0.36)	0.05(0.37)	-0.00(0.37)	0.01(0.37)	-0.03(0.38)	-0.01(0.39)	-0.00(0.38)	-0.01(0.38)	-0.00(0.38)
126	0.51(0.46)	0.58(0.46)	0.56(0.49)	0.46(0.45)	0.48(0.47)	0.47(0.47)	0.46(0.46)	0.49(0.48)	0.46(0.47)	0.47(0.49)
ALZ Blk	—	—	—	—	—	0.19(0.69)	0.35(0.69)	0.34(0.77)	0.20(0.70)	0.20(0.71)
122	0.42(0.17)	0.43(0.18)	0.43(0.19)	0.40(0.18)	0.42(0.17)	0.44(0.18)	0.45(0.18)	0.45(0.18)	0.45(0.18)	0.44(0.18)
ALL	0.24(0.23)	0.49(0.24)	0.44(0.34)	0.02(0.08)	0.16(0.19)	0.06(0.31)	0.56(0.33)	0.50(0.89)	-0.00(0.07)	0.04(0.23)

Table E.3: Estimated relative risks under Poisson models.

Appendix F
EM/Gibbs/Importance Sampling Steps

1. Select a starting point $\theta^{(0)} = (\alpha^{(0)}, \lambda^{(0)})'$.

2. **for** iteration $s = 0, 1, 2...$, **do** (EM Steps)

   - 1).Generate k clusters $(\mathbf{O}, \mathbf{r})^{(s,0)}$ from $\lambda^{(s)}$ to satisfy $\{\delta = k\}$.
   - 2). **for** $i = 1, 2, ...I$, **do** (I sets of Gibbs samples)
     - Initialize $(\mathbf{O}, \mathbf{r})^{(s,i)}$ by its previous version $(\mathbf{O}, \mathbf{r})^{(s,i-1)}$.
     - **for** $j = 1, 2, ..., k$, **do** (generate a set of clusters from (5.8))
       - **for** $l = 1, 2, ...L$, **do** (importance sampling for jth cluster)
         - a). Generate $(\mathbf{o_j}, r_j)^*$ from $\varphi_{\lambda_o^{(s)}}(\mathbf{o_j})\varphi_{\lambda_r^{(s)}}(r_j)$. Replace the jth cluster in $(\mathbf{O}, \mathbf{r})^{(s,i)}$ and denote the new set as $(\mathbf{O}, \mathbf{r})^*$: $((\mathbf{o}_1, r_1)^{(s,i)}, ..., (\mathbf{o}_{j-1}, r_{j-1})^{(s,i)}), (\mathbf{o_j}, r_j)^*, (\mathbf{o}_{j+1}, r_{j+1})^{(s,i)}, ..., (\mathbf{o}_k, r_k)^{(s,i)})$.
         - b). Check if $(\mathbf{O}, \mathbf{r})^*$ makes $\{\delta = k\}$. If no, go back to a).
         - c). Denote the generated cluster as $(\mathbf{o_j}, r_j)^{[l]}$. Calculate weight $w_l = \frac{\prod_{j \neq j'} [g_2((\alpha_j)^{(s)}, (\alpha_{j'})^{(s)})]^{Z_{j,j'}(\mathbf{y}, (\mathbf{O}, \mathbf{r})^*)}}{[c(\mathbf{y}, (\mathbf{O}, \mathbf{r})^*, (\theta)^{(s)})]^n} (\alpha_j^{(s)})^{Z_j(\mathbf{y}, (\mathbf{O}, \mathbf{r})^*)}$.

         end for (*end loop* l)
       - Select one set from the L samples $(\mathbf{o_j}, r_j)^{[1]}$,....,$(\mathbf{o_j}, r_j)^{[L]}$ with respective probabilites $(p_1, p_2, ...p_L)$ where $p_l = w_l / \sum_{k=1}^{L} w_k$. Use this cluster to update the jth cluster in $(\mathbf{O}, r)^{(s,i)}$.

     end for (*end loop* j)

   end for (*end loop* i)

- 3).(E-step) Approximate the conditional expectations $Q(\theta|\theta^{(s)}) = Q_1(\alpha|\theta^{(s)}) + Q_2(\lambda|\theta^{(s)})$ at the current $\theta^{(s)}$ by Monte-Carlo simulation (after m burn-ins): (Note: $\theta^{(s)}$ is intrinsic in generating Gibbs samples $(\mathbf{O}, \mathbf{r})^{(s,i)}$)).

$$Q_1(\alpha|\theta^{(s)}) \approx \frac{1}{M-m} \sum_{i=m+1}^{M} \{\sum_{j=1}^{k} Z_j(\mathbf{y}, (\mathbf{O}, \mathbf{r})^{(s,i)}) log\alpha_j + \sum_{j \neq j'} Z_{j,j'}(\mathbf{y}, (\mathbf{O}, \mathbf{r})^{(s,i)}) log[g_2(\alpha_j, \alpha'_j)] - nlog[c(\mathbf{y}, (\mathbf{O}, \mathbf{r})^{(s,i)}, \alpha)]\}$$

$$Q_2(\lambda|\theta^{(s)}) \approx \frac{1}{M-m} \sum_{i=m+1}^{M} \sum_{j=1}^{k} \left\{log[\varphi_{\lambda_o}(\mathbf{o_j})^{(s,i)} \times \varphi_{\lambda_r}(r_j)^{(s,i)}]\right\} \quad \text{(F.1)}$$

- 4). (M-step) maximize Q_1 and Q_2 to get $\theta^{(s+1)}$.

3. **end for** if $\|\theta^{(s+1)} - \theta^{(s)}\|$ is very small. Suppose the last step has S for the index s, then the cluster $(\mathbf{o_j}, r_j)$ is estimated as
$\hat{\mathbf{o_j}} = \frac{1}{M-m} \sum_{i=m+1}^{M} \mathbf{o_j}^{(S,i)}$, $\hat{r_j} = \frac{1}{M-m} \sum_{i=m+1}^{M} r_j^{(S,i)}$

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716-723.

[2] Akyildiz, I., Su, W., Sankarasubramaniam, Y. E., and Cayiric, E. (2002a). Survey on sensor networks. *IEEE Communications Magazine, August 2002*, 102-114.

[3] Akyildiz, I., Su, W., Sankarasubramaniam, Y. E., and Cayiric, E. (2002b). Wireless sensor networks: a survey. *Computer Networks*, 38:393-422.

[4] Antman, E.M., Bennett, J.S., Daugherty, A., Furberg, C., Roberts, H., and Taubert, K.A. (2007). Use of nonsteroidal antiinflammatory drugs. *Circulation*, 115:1634-1642.

[5] Balakrishnan, N. and Koutras, M. (2001). *Runs and Scans with Applications*, John Wiley and Sons.

[6] Bollen, C., Uiterwaal, C., van Vught, A., and van der Tweel, I. (2006). Sequential meta-analysis of past clinical trials to determine the use of a new trial. *Epidemiology*, 17:644-649.

[7] Carpenter, J., Pocock,S., and Lamm, C.J. (2002). Coping with missing data in clinical trials: A model-based approach applied to asthma trials. *Statistics in Medicine* 21:1043-1066.

[8] Carpenter, T., Cheng, J., Roberts, F. and Xie, M. (2010). Sensor Management Problems of Nuclear Detection. (Invited Book Chapter) *Safety and Risk Modeling and Their Applications*, Springer.

[9] Cheng, J. and Madigan, D. (2009). Bayesian approaches to aspects of the Vioxx trials: non-ignorable dropout and sequential meta-analysis. *The Oxford Handbook of Applied Bayesian Analysis*, Oxford University Press.

[10] Cheng, J. and Xie, M. (2009). A latent model to detect multiple spatial clusters with application in a mobile sensor network for the surveillance of nuclear materials. *Proceedings of Joint Statistical Meeting, 2009*.

[11] Cleveland, W.S., Devlin, S.J., and Grosse, E. (1988). Regression by local fitting. *Journal of Econometrics*, 37:87114.

[12] Cleveland, W.S. and Grosse, E. (1991). Computational methods for local regression. *Statistics and Computing*, 1:4762.

[13] The Cochrane Library (2007). Wiley:Chichester, Issue 2.

[14] Cooper, G. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9:309-347.

[15] Demattei, C., Molinari, N., and Daures, J. (2006). Spatclas: an R package for arbitrarily shaped multiple spatial cluster detection for case event data. *Computer Methods and Programs in Biomedicine*, 84:42-49.

[16] Demattei, C., Molinari, N., and Daures, J. (2007). Arbitrary shaped multiple spatial cluster detection for case event data. *Computational Statistics and Data Analysis*, 51:3931-3945.

[17] Dembo, A and Karlin, S. (1992). Poisson approximation for r-scan processes. *Annals of Applied Probability*, 2:329357.

[18] Demirtas, H. and Schafter, J.L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* 22:2553-2575.

[19] Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, Series B (Methodological), 39(1):1-38.

[20] Denison, D. and Holmes, C. (2001). Bayesian partitioning for estimating disease risk. *Biometrics*, 57:143-149.

[21] DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7:177-188.

[22] Federal Emergency Management Agency (FEMA). (2008). Are You Ready? *http://www.fema.gov/areyouready/nuclear_blast.shtm.*

[23] Fu, J.C. and Lou, W.Y. (2003). *Distribution theory of runs and patterns and its applications: a finite Markov Chain embedding approach*, World Scientific.

[24] Gangnon, R. and Clayton, M. (2000). Bayesian detection and modeling of spatial disease maps. *Biometrics*, 56:922-935.

[25] Gangnon, R. and Clayton, M. (2003). A hierarchical model for spatially clustered disease rates. *Statistics in Medicine*, 22:3213-3228.

[26] Gangnon, R. and Clayton, M. (2004). Likelihood-based tests for localized spatial clustering of disease. *Environmetrics*, 15:797-810.

[27] Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC.

[28] Glaz, J. and Balakrishnan, N. (1999). *Scan Statistics and Applications.*, Boston: Birkhauser.

[29] Glaz, J., Naus, J., and Wallenstein, S. (2001) *Scan Statistics and Applications*, New York:Springer.

[30] Henderson, W.G., Moritz, T., Goldman, S., and Copeland, J. (1995). Use of cumulative meta-analysis in the design, monitoring, and final analysis of a clinical trial: a case study. *Controlled Clinical Trials*, 16:331-341.

[31] Heckerman, D. (1996). A tutorial on learning with Bayesian networks. *Technical Report MSR-TR-95-06*, Microsoft Research.

[32] Higgins, J. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*, 15:2733-2749.

[33] Hochbaum, D. (2008). The multi-sensor nuclear threat detection problem. *Proceedings of the Eleventh INFORMS Computing Society (ICS) Conference, 2008*, in press.

[34] Imbens, G.W. and Rubin, D.B. (1997a). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, 25:305-327.

[35] Imbens, G.W. and Rubin, D.B. (1997b). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economics Studies*, 64:555-574.

[36] Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90:773-795.

[37] Knorr-Held, L. and RaBer, G. (2000) Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56:13-21.

[38] Konstam, M.A., Weir, M.R., Reicin, A., Shapiro, D., Sperling, R.S., Barr, E., and Gertz, B. (2001). Cardiovascular thrombotic events in controlled clinical trials of roefecoxib. *Circulation*, 104:2280-2288.

[39] Kulldorff, M. and Nagarwalla, N. (1995). Spatial disease clusters: Detection and infection. *Statistics in Medicine*, 14:799-810.

[40] Lawson, A. (1995). Markov Chain Monte Carlo methods for putative pollution source problems. *Environmental Epidemiology*, 14:2473-2486.

[41] Lewis, H.G. (1963). *Unionism and relative wages in the United States: an empirical inquiry*. Chicago, IL: University of Chicago Press.

[42] Little, R.J.A. (1994). A class of pattern-mixture models for normal missing data. *Biometrika* 81:471-483.

[43] Little, R.J.A. and Wang, Y (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrika* 52:98-111.

[44] Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: Wiley.

[45] Little, R.J.A. and An, H. (2004). Robust likelihood-based analysis of multivariate data with missing values. *Statistica Sinica*, 14:949-968.

[46] Madigan, D. (2005). Bayesian data mining for health surveillance. *Spatial and Syndromic Surveillance for Public Health*. John Wiley & Sons, Ltd.

[47] Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63:215-232.

[48] McClellan, M. (2007). Drug safety reform at the FDA - pendulum swing or systematic improvement? *The New England Journal of Medicine*, 356:1700-1702.

[49] Molinari, N., Bonaldi, C., and Daures, J. (2001). Multiple temporal cluster detection. *Biometrics*, 57:577-583.

[50] Naus, J. and Wallenstein, S. (2004). Multiple window and cluster size scan procedures. *Methodology and Computing in Applied Probability*, 6:389400.

[51] Neill, D.B., Moore, A.W.,and Cooper G.F. (2006). A Bayesian spatial scan statistic. *Advances in Neural Information Processing Systems*, 18:1003-1010.

[52] Normand, S.L.T. (1999). Meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18(3):321-359.

[53] Purdue University. (2008). Cell phone sensors detect radiation to thwart nuclear terrorism.
http://news.uns.purdue.edu/x/2008a/080122FischbachNuclear.html

[54] Reicin, A.S., Shapiro, D., Sperling, R.S., Barr, E., and Yu, Q. (2002). Comparison of cardiovascular thrombotic events in patients with osteoarthritis treated with rofecoxib versus nonselective nonsteroidal anti-inflammatory drugs. *American Journal of Cardiology*, 89:204-209.

[55] Ridgeway, G. and Madigan, D. (2002). Bayesian analysis of massive datasets via particle filters. *Proceedings of the eighth ACM SIGKDD*. Statistical Methods 5-13.

[56] Rothman, K.J., Greenland, S., and Lash, T. (2008). *Modern Epidemiology*. Lippincott Williams & Wilkins.

[57] Rubin, D.B. (1976). Inference and missing data. *Biometrika* 62,581-592

[58] Rubin, D.B. (1978). Bayesian inference for causal effects. *Annals of Statistics*, 6:34-58.

[59] Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley.

[60] Rubin, D.B. (1990). Formal modes of statistical inference for causal effects: the role of randomization. *Journal of Statistical Planning and Inference*, 25:279-292.

[61] Rubin, D.B. (1996). Multiple imputation after +18 years (with discussion). *Journal of American Statistical Association*, 91:473-489.

[62] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: CRC Press/Chapman & Hall.

[63] Schafer, J. L. (1997). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8:3-15.

[64] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461-464.

[65] Spiegelhalter, D.J. and Lauritzen, S.L. (1990). Sequential updating of conditional probabilities on direct graphical structures. *Networks*, 20:579-605.

[66] Spiegelhalter, D.J., Freedman, L., and Parmar, M. (1994). Bayesian analysis of randomized trials (with discussion). *Journal of the Royal Statistical Society (Series B)*, 157, 357–416.

[67] Spiegelhalter, D.J., Thomas, A., Best, N.G., and Gilks, W.R. (1995). *BUGS: Bayesian Inference using Gibbs sampling, Version 0.50.* MRC Biostatistics Unit, Cambridge.

[68] Spiegelhalter, D., Best, N., Carlin, B., and Van de Linde, A. (2002) Bayesian measures of complexity and fit (with discussion). *Journal of the Royal Statistical Society* (Series B), 64, 583-640.

[69] Su, X., Wallenstein, S. and Bishop, D. (2001). Non-overlapping clusters: approximation distribution and application to molecular biology. *Biometrics*, 57:420-426.

[70] Sun, Q. L. (2008). Statistical modeling and inference for multiple temporal or spatial cluster detection. *Ph.D. thesis*, Department of Statistics, Rutgers University.

[71] Sutton, A.J. and Higgins, J.P.T. (2007). Recent developments in meta-analysis. *Statistics in Medicine*, 27:625-650.

[72] Sweeting, M.J., Sutton, A.J., and Lambert, P.C. (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*, 23:1351-1375.

[73] Wallenstain, S. and Naus, J. (2004). Scan statistics for temporal surveillance for biological terrorism. *Morbidity & Mortality Weekly Report*, 53:74-78.

[74] Wein, L., Wilkins, A., Baveja, M., and Flynn, S. (2006). Preventing the importation of illicit nuclear materials in shipping containers. *Risk Analysis*, vol. 26.

[75] Weir, M.R., Sperling, R.S., Reicin, A., and Gertz, B.G. (2003). Selective COX-2 inhibition and cardiovascular effects: A review of the rofecoxib development program. *American Heart Journal*, 146:591-604.

[76] Whitehead, A. and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomised clinical trials. *Statistics in Medicine*, 10:1665-1677.

[77] Wood, A.M., White, I.R., Hillsdon, M., and Carpenter, J. (2004). Comparison of imputation and modeling methods in the analysis of a physical activity trial with missing outcomes. *International Journal of Epidemiology*, 34:89-99.

[78] Xie, M., Sun, Q., and Naus, J. (2009). A latent model to detect multiple temporal clusters. *Biometrics*, 65:1011-1020.

[79] Yau, L.H.Y. and Little, R.J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, 96(456):1232-1244.

Vita

Jerry Q. Cheng

1989 B.S. in Physics, Fudan University.

1991 M.S. in Physics, Virginia Tech.

1994 M.S. in Statistics, North Carolina State University.

1999 M.S. in Computer Sciences, North Carolina State Univeristy.

2010 PhD. in Statistics, Rutgers University.

1990 - 1991 Teaching Assistant, Department of Physics, Virginia Tech, Blacksburg, VA.

1991 - 1992 Teaching Assistant, Department of Physics, North Carolina State University, Raleigh, NC.

1992 - 1994 Teaching Assistant, Department of Statistics, North Carolina State University, Raleigh, NC.

1994 - 1998 Statistician, Mathematica Policy Research Inc., Plainsboro, NJ.

1998 - 2000 Technical Staff Member, AT&T Labs, Middletown, NJ.

2000 - 2002 Senior Software Engineer, Tellium Inc., Oceanport, NJ.

2002 - 2006 Technical Staff Member, AT&T Labs, Middletown, NJ.

2007 Instructor, Department of Statistics, Rutgers University, New Brunswick, NJ.

2007 - 2010 Research Assistant, DIMACS, Rutgers Univsersity, New Brunswick, NJ.

Selected Publications

2009 Jerry Cheng and David Madigan. Bayesian Approaches to Aspects of the Vioxx Trials: Non-ignorable Dropout and Sequential Meta-Analysis. Handbook of Applied Bayesian Statistics, 2008.

2009 Jerry Cheng and Minge Xie. A Latent Model to Detect Multiple Spatial Clusters with Application in a Mobile Sensor. In Proceedings of Joint Statistical Meetings (JSM), 2009.

2009 Jerry Cheng, Minge Xie and Fred Roberts. Design and Deployment of a Mobile Sensor Network in Surveillance of Nuclear Materials in Metropolitan Areas. In Proceedings of the 15th ISSAT International Conference on Reliability and Quality in Design, 2009.

2010 Tammi Carpenter, Jerry Cheng, Fred Roberts and Minge Xie. Sensor Management Problems of Nuclear Detection. Invited Book Chapter, Safety and Risk Modeling and Their Applications, Springer, 2010