

ABSTRACT

Bayesian Evaluation of Surrogate Endpoints

Chunyao Feng

Mentor: John W. Seaman Jr., Ph.D.

To save time and reduce the size and cost of clinical trials, surrogate endpoints are frequently measured instead of true endpoints. The proportion of the treatment effect explained by surrogate endpoints (*PTE*) is a widely used, albeit controversial, validation criteria. Frequentist and Bayesian methods have been developed to facilitate such validation. The former does not formally incorporate prior information; a critical issue since confidence intervals on *PTE* is often unacceptably wide. Both the Bayesian and frequentist approaches may yield estimates of *PTE* outside the unit interval. Furthermore, the existing Bayesian method offers no insight into the prior used for *PTE*, making prior-to-posterior sensitivity analyses problematic. We proposed a fully Bayesian approach that avoids both of these problems. We also consider the effect of interaction on inference for *PTE*. As an alternative to the use of *PTE*, we develop a Bayesian model for relative effect and the association between surrogate and true endpoints, making use of power priors.

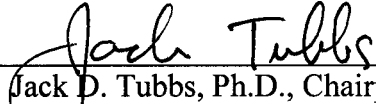
Bayesian Evaluation of Surrogate Endpoints

by

Chunyao Feng

A Dissertation

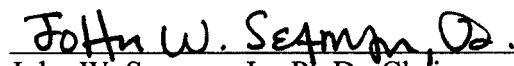
Approved by Department of Statistical Science



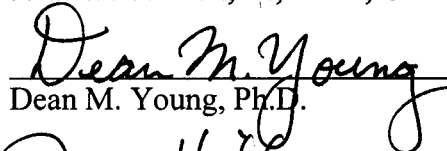
Jack D. Tubbs, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

Approved by Dissertation Committee



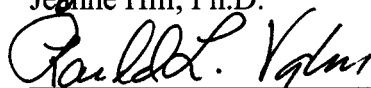
John W. Seaman, Jr., Ph.D., Chairman



Dean M. Young, Ph.D.



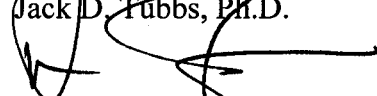
Jeanne Hill, Ph.D.



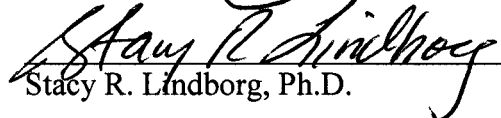
Randal L. Vaughn, Ph.D.



Jack D. Tubbs, Ph.D.

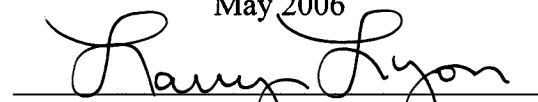


Dennis A. Johnston, Ph.D.



Stacy R. Lindborg, Ph.D.

Accepted by the Graduate School
May 2006



J. Larry Lyon, Ph.D., Dean

Copyright © 2006 by Chunyao Feng

All rights reserved

TABLE OF CONTENTS

LIST OF FIGURE.....	vii
LIST OF TABLES	ix
ACKNOWLEDGMENTS	xi
CHAPTER ONE Basic Concepts in Surrogate Endpoints Evaluation	1
1.1. Introduction to Surrogate Endpoints.....	1
1.1.1 Why Use a Surrogate Endpoint.....	2
1.1.2 A Valid Surrogate Endpoint.....	3
1.2 <i>PTE</i> , <i>RE</i> and ρ	5
1.2.1 Definitions.....	5
1.2.2 Fieller's Theorem to Calculate Confidence Limits for <i>PTE</i> and <i>RE</i>	6
1.3 Advantages of Bayesian Approaches in Surrogate Endpoints Evaluation	7
CHAPTER TWO Bayesian Evaluation of Surrogate Endpoints Using <i>PTE</i>	9
2.1 Introduction to <i>PTE</i>	9
2.1.1 Inferences for <i>PTE</i>	10
2.1.2 Maximum Likelihood Estimation of <i>PTE</i>	10
2.1.3 Cowles' Bayesian Method for Estimation of <i>PTE</i>	12
2.1.4 Disadvantages of <i>PTE</i>	13
2.2 Conditional Prior Approach.....	14
2.2.1 Assumptions	15

2.2.2 Conditional Prior Structure for β and β_γ	15
2.3 A Small Simulation Study	16
2.3.1 Data Generation	16
2.3.2 Prior Structure for β and β_γ	18
2.3.3 Simulation Results.....	19
2.4 Induced Prior Structure for <i>PTE</i>	22
2.4.1 Induced Prior for <i>PTE</i> — Ratio of Two Dependent Beta Distributions ($\beta \geq \beta_\gamma \geq 0$)	22
2.4.2 Induced Prior for <i>PTE</i> — Ratio of Two Dependent Beta Distributions ($\beta \leq \beta_\gamma \leq 0$)	25
2.4.3 Example Using Conditional Beta Priors and the Induced Prior on <i>PTE</i>	26
2.4.4 Induced Prior for <i>PTE</i> — Ratio of Two Independent Normal Distributions	27
2.5 Support For Uniform Prior.....	29
CHAPTER THREE Simulation Study.....	30
3.1 Introduction.....	31
3.2 Simulation Study.....	32
3.2.1 Prior Structures.....	32
3.2.2 Small Sample Size ($N = 40$)	34
3.2.3 Moderate Sample Size ($N = 200$)	39
3.2.4 Conclusions.....	43
3.3 Informative Prior from Historical Data.....	44
3.3.1 Posterior Based on Historical Data.....	44
3.3.2 Prior Structures Based on Historical Data (sample size $N_h = 20$ and $N_h = 100$).....	45

3.4 Sensitivity Analysis	51
CHAPTER FOUR Modeling Interaction Effect	55
4.1 The Effect of Assuming Interaction Is Absent When It Is Not.....	55
4.2 The Influence of Interactions on <i>PTE</i> Estimation	57
4.3 The Affect of Interaction on <i>PTE</i> Estimation	61
4.3.1 Scenario 1: Small Interaction ($\eta = 0.2$)	63
4.3.2 Scenario 2 ($\eta = 1$)	66
4.3.3 Scenario 3 ($\eta = 1.5$)	67
4.4 Estimating <i>PTE</i> When Small Interaction is Present	69
4.5 Joint Distribution of <i>PTE</i> and η	71
4.5.1 Introduction.....	71
4.5.2 An Example	73
4.6 Bayesian Model Comparisons	74
4.6.1 Overview of Model Comparison	74
4.6.2 Bayesian Model Comparison Methods.....	75
4.6.3 Conclusion	78
CHAPTER FIVE Relative Effect	79
5.1 Introduction.....	79
5.2 Inferences for <i>RE</i>	81
5.2.1 Frequentist Inference for <i>RE</i> and ρ	81
5.2.2 Bayesian Inference for <i>RE</i> and ρ for Longitudinal Data.....	82
5.2.3 Bayesian Inference for <i>RE</i> and ρ for Normal Endpoints	83

5.3 Power Priors	84
5.4 An Example	85
5.4.1 Data Generation	85
5.4.2 Prior Structure	87
5.4.3 Simulation Results for Power Prior	90
5.4.4 Comparison of Posteriors under Power, Informative, and Diffuse Priors	93
5.5 Power Prior for Bayesian Estimation for <i>PTE</i>	95
5.5.1 Bayesian Estimation for <i>PTE</i> with a Power Prior	95
5.5.2 Posterior Results under the Power Prior	97
CHAPTER SIX Conclusions	99
APPENDICES	101
APPENDIX A Convergence Issues in MCMC Simulations	102
APPENDIX B SAS IML Program to Generate Data	105
APPENDIX C SAS Procedure to Perform Frequentist Analysis	107
APPENDIX D R Programs	108
APPENDIX E WinBUGS Programs	111
REFERENCES	115

LIST OF FIGURES

Figure 1.	Why strong correlation is not sufficient for a valid surrogate endpoint....	4
Figure 2.	Posterior distribution of <i>PTE</i> using diffuse normal prior structure	20
Figure 3.	Histograms of the confidence interval widths	22
Figure 4.	Posterior of <i>PTE</i> vs. Prior of <i>PTE</i>	27
Figure 5.	Induced prior for <i>PTE</i> using diffuse normal prior for β and β_γ	28
Figure 6.	Posterior <i>PTE</i> and upper support for d in $\beta \sim \text{beta}_{[0, d]}(a, b)$	29
Figure 7.	Induced priors for <i>PTE</i>	34
Figure 8.	Posterior distributions for <i>PTE</i>	36
Figure 9.	Induced informative priors for <i>PTE</i>	48
Figure 10.	Induced <i>PTE</i> Priors using Conditional Beta Priors in (3.10)	52
Figure 11.	Posterior density of <i>PTE</i>	53
Figure 12.	Posterior density of <i>PTE</i> with four different β_γ priors and one fixed β prior structure.....	54
Figure 13.	Posterior density for η	60
Figure 14.	Median and 95% credible set values for $h(S)$	61
Figure 15.	Induced prior for <i>PTE</i> with four prior structures.....	63
Figure 16.	Posterior densities for <i>PTE</i> and η with sample size 40 and 200 using prior 1	65
Figure 17.	Posterior densities for the interaction using three prior structures for $N=40$	65
Figure 18.	Posterior densities for η with samples size 40 and 200 using prior 1. The true value is $\eta = 1$	67

Figure 19. Posterior densities of η for samples with size 40 and 200 when true $\eta = 1.5$	68
Figure 20. Posterior densities for β_γ and β'_γ	70
Figure 21. Posterior densities for β and η	70
Figure 22. Median and 95% credible set values for $h(S)$	71
Figure 23. Marginal probabilities for PTE and $ \eta $	72
Figure 24. Posterior distributions for PTE and η	74
Figure 25. Posterior distributions for RE and ρ	92
Figure 26. Posterior distributions for RE with different a_0	92
Figure 27. Posterior distributions for RE with different prior structures.....	94
Figure 28. Posterior distributions for RE with different prior structures.....	95
Figure 29. Posterior distribution for PTE with power prior.....	97
Figure A.1 Autocorrelation for one chain of posterior PTE distribution before thin the sequences	103
Figure A.2 Autocorrelation for one chain of posterior PTE distribution after thin the sequences.....	104

LIST OF TABLES

Table 1. Simulation results comparison.....	21
Table 2. Prior structures for non-informative prior simulation.....	33
Table 3. Simulation results with four prior structures when true $PTE = 0.5$ and $N = 40$	36
Table 4. Simulation results with four prior structures when true $PTE = 0.7$ and $N = 40$	37
Table 5. Simulation results with four prior structures when true $PTE = 0.9$ and $N = 40$	39
Table 6. Simulation results with four prior structures when true $PTE = 0.5$ and $N = 200$	40
Table 7. Simulation results with four prior structures when true $PTE = 0.7$ and $N = 200$	41
Table 8. Simulation results with four prior structures when true $PTE = 0.9$ and $N = 200$	42
Table 9. Prior structures and 95% intervals for parameters.....	46
Table 10. Normal prior structures for β , β_γ and their 95% intervals	47
Table 11. Conditional beta prior structures for β , β_γ and their 95% intervals	47
Table 12. Induced prior structures for PTE and their 95% Intervals	48
Table 13. Posterior statistics for PTE with informative prior structures	49
Table 14. Posterior credible sets for PTE using historical data for conditional beta priors compared to the conditional uniform priors	50
Table 15. Summary of induced PTE priors using the conditional Beta Priors	52
Table 16. Posterior median and 95% credible set for PTE for different combination of prior for β and β_γ with sample size 40	53

Table 17. Posterior median <i>PTE</i> and 95% credible set for <i>PTE</i> for different combination of prior for β and β_γ with sample size 200	54
Table 18. Estimators for <i>PTE</i> and η when true $\eta = 0.2$	64
Table 19. Estimators for <i>PTE</i> and η when true $\eta = 1$	66
Table 20. Estimators for <i>PTE</i> and η when true $\eta = 1.5$	68
Table 21. Estimators for <i>PTE</i> and η when true $\eta = 0.2$	77
Table 22. Model comparison statistics for scenario with $\eta=1.5$	78
Table 23. Initial prior distribution in power prior distribution	88
Table 24. Informative normal prior distribution	89
Table 25. Diffuse normal prior distribution.....	90
Table 26. Posterior and MLE estimations for <i>RE</i> , ρ , α , β and Σ	91
Table 27. Power prior with informative Wishart structure	93
Table 28. Posterior mean and 95% credible sets for <i>RE</i> , α , β and ρ	93
Table 29. Initial prior distributions in power prior method	96

ACKNOWLEDGMENTS

I would like to express my sincerest and deepest appreciation to my advisor, Dr. John Seaman for his guidance, support, encouragement and patience throughout the entire course of this research. Dr. John Seaman is one of the best professors I ever met and his diligence and his passion for his teaching and research clearly set up a role model for me to follow in my future research career.

My genuine gratitude also extends to Dr. Stacy Lindborg for her valuable academic advices and her generous supports in the past two years.

I would like to acknowledge all of my committee members: Dr. Stacy Lindborg, Dr. Jack Tubbs, Dr. Dean Young, Dr. Dennis Johnston, Dr. Jeanne Hill and Dr. Randal Vaughn. Thank all of them for their time and their kind assistances.

I am truly grateful to the faculty, staff members and all my classmates in Department of Statistics at Baylor University. I especially thank Dr. Tom Bratcher and Dr. Jack Tubbs for their accepting me into our Ph.D. program in Statistics. I had a wonderful educational experience at Baylor and I really enjoyed my three-year graduate study at Waco.

Finally, I would like to thank my husband, Huimin Bie, for his continuous love and unselfish supports all the time. I also wish to express my appreciation to my parents and my brother for their love and everything they have done for me throughout my life.

CHAPTER ONE

Basic Concepts in Surrogate Endpoints Evaluation

The purpose of this dissertation is to introduce new Bayesian approaches for evaluating surrogate endpoints. In this chapter, we review basic concepts regarding surrogate endpoint evaluation and common issues in the validation process. We develop a new Bayesian approach, using conditional priors, to estimate PTE (the proportion of treatment effects explained by surrogate endpoints) in Chapter 2. We shall study the basic properties of the conditional prior method and compare it with currently available methods. In Chapter 3, simulation studies are conducted to further explore properties of the conditional prior approach. Prior-to-posterior sensitivity analyses are performed as well. In Chapter 4, we study the effect of interaction on the estimation of PTE . There we make use of Bayesian model comparison techniques. Finally, in Chapter 5, we explore the Bayesian analysis of relative effect (RE) and association (ρ) between surrogate endpoints and true endpoints after adjustment for treatment. They are considered as alternatives to PTE . For making use of historical data we derive power prior structures for RE and ρ . Examples are given to illustrate Bayesian inferences for PTE , RE and ρ using the power prior structure.

1.1. Introduction to Surrogate Endpoints

Clinical trials designed to demonstrate an endpoint such as mortality reduction are typically long and costly. Surrogate endpoints are outcomes that reflect an ultimate endpoint of clinical interest but which are observable earlier. They may or may not be of

interest in themselves. For example, reductions in cholesterol are known to be associated with mortality benefits in cardiac health. Blood pressure as a risk factor in strokes and heart attacks is another example. Bone mineral density is used as a surrogate for bone fracture risk. CD-4 lymphocyte levels serve as a surrogate endpoint in trials concerning HIV infection.

1.1.1 Why Use a Surrogate Endpoint

Molenberghs, Burzykowski, and Alonso (2004) provided definitions of true endpoints, biomarkers, and surrogate endpoints. A true endpoint is “a characteristic or variable that reflects how a patient feels, functions, or survives”. A biomarker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”. The U.S. Food and Drug Administration (FDA) defined a surrogate endpoint as “a laboratory measurement or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy”. Therefore, a surrogate endpoint is usually viewed as a biomarker. A biomarker, however, is not necessarily a surrogate endpoint unless it is “validated”. How to validate a biomarker remains a controversial subject.

Molenberghs *et al.* (2004) summarized the following motivations for the use of surrogate points:

- Surrogate endpoints may be easier and more convenient to measure;
- Surrogate endpoints can be observed more frequently;

- Surrogate endpoints are less subject to competing risks and less affected by other treatments;
- Surrogate endpoints can improve the early decision-making process and reduce the time to assess clinically important outcomes.

Despite the attractive benefits of using surrogate points in clinical trials, researchers still have to face some major challenges:

- Do surrogate endpoints reflect the real clinical outcomes?
- Is it feasible to obtain precise and reliable surrogate endpoints?
- Will the study of surrogate endpoints be possible if the evaluation procedures are not acceptable to the patients?

These issues have a bearing on the validity of a surrogate endpoint.

1.1.2 A Valid Surrogate Endpoint

Use of a surrogate endpoint can reduce the size and length of clinical trials. Despite this potential, there is no widely accepted agreement about what constitutes a valid surrogate endpoint. Prentice (1989) provided two sufficient conditions for a valid surrogate endpoint. First, a valid surrogate endpoint needs to have a powerful and consistent association with the clinical endpoint of interest. This requires that the changes in surrogate endpoints correlate with changes in the true endpoints. This is necessary but not sufficient. Indeed, it is one of the most common misunderstandings that strong correlation between surrogate variables and clinical variables is adequate to validate a surrogate. Fleming and DeMets (1996) illustrated this with a graphic similar to that in Figure 1.

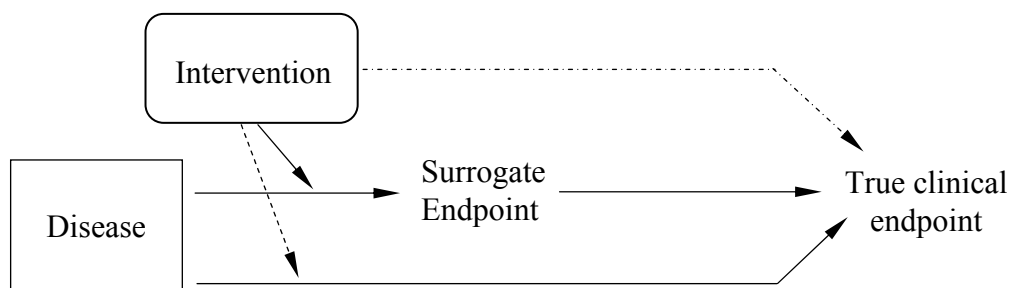


Figure 1. Why strong correlation is not sufficient for a valid surrogate endpoint. Here ----- indicates an unintended beneficial effect and - - - - - indicates an unintended negative effect. Adapted from Fleming & DeMets (1996).

Suppose the disease process influences the risk of the true endpoints through multiple pathways. If the proposed surrogate endpoint lies in only one of the pathways and the intervention does not actually affect all the pathways, the effect of treatment on the true endpoint could be over- or under-estimated by relying on the effect of the proposed surrogate.

The second requirement for a valid surrogate endpoint is that it must fully capture the net effect of the intervention on the true endpoint. This requirement is much more difficult to satisfy and verify than the first requirement.

It is not feasible to prove that a marker is a valid surrogate by conducting statistical analyses to verify Prentice's (1989) criteria. Therefore, to interpret these two requirements, Prentice provided four operational criteria in the same paper:

1. The surrogate endpoint has to be in the pathway and capture the effect of treatment;
2. The treatment has to have a biological effect on the true endpoint;
3. The surrogate has a significant impact on the true endpoint;
4. The full treatment effect has to be reflected by surrogate.

Prentice (1989) used the following regression equations to model these four criteria:

$$\mathcal{E}(S|T) = \mu' + \alpha T \quad (1.1)$$

$$\mathcal{E}(E|T) = \mu + \beta T \quad (1.2)$$

$$\mathcal{E}(E|S) = \mu_E + \gamma_E S \quad (1.3)$$

$$\mathcal{E}(E|T, S) = \mu_\gamma + \beta_\gamma T + \gamma S \quad (1.4)$$

where $\mathcal{E}(\cdot)$ denotes conditional expectation, T is used to denote the treatment effect, S represents surrogate endpoint and E means true or “final” endpoint.

The effect of the treatment on the surrogate is modeled by (1.1) and the parameter of interest is α . Similarly, the effect of the treatment on the true endpoint is reflected by β in (1.2). The effect of the surrogate on the true endpoint is represented by γ_E . The first three operational criteria require testing the significance of three parameters, α , β and γ_E . The last criterion requires that β_γ be non-significant. This raises a conceptual difficulty.

1.2 PTE, RE and ρ

1.2.1 Definitions

Freedman, Graubard, and Schatzkin (1992) argued that the non-significance of β_γ does not prove that the effect of treatment upon the true endpoint is fully captured by the surrogate. Therefore, he proposed a new criterion to evaluate a surrogate endpoint: the proportion of treatment effect explained by surrogate endpoint (*PTE*). *PTE* is defined as

$$PTE = \frac{\beta - \beta_\gamma}{\beta} = 1 - \frac{\beta_\gamma}{\beta} \quad (1.5)$$

where β and β_γ are from (1.2) and (1.4).

They can be obtained by fitting these equations simultaneously. A perfect surrogate endpoint can account for all mechanisms of actions. With $PTE < 1$, the surrogate endpoint explains only part of the treatment effect on the true endpoint. A “good” surrogate is one that explains a large proportion of the treatment’s effect on the true endpoint.

Buyse and Molenberghs (1998) proposed alternative criteria to PTE for the evaluation of surrogate endpoints. Suppose S and E follow a bivariate normal distribution with means (1.1) and (1.2) and variance-covariance structure

$$\begin{bmatrix} \sigma_{SS}^2 & \sigma_{SE} \\ \sigma_{SE} & \sigma_{EE}^2 \end{bmatrix}.$$

Define the relative effect (RE) to be

$$RE = \frac{\beta}{\alpha}.$$

RE evaluates “the effect of treatment T on the true endpoint E relative to that of treatment effect on surrogate endpoints”. Additionally, “the association between surrogate endpoint and true endpoint after adjustment for treatment” is the correlation given by

$$\rho = \frac{\sigma_{SE}}{\sqrt{\sigma_{SS}^2 \sigma_{EE}^2}}.$$

A joint model for S and E given T affords simultaneous estimation of RE and ρ .

1.2.2 Fieller’s Theorem to Calculate Confidence Limits for PTE and RE

Maximum likelihood methods can be used to estimate PTE , RE and ρ . Fieller’s theorem or the delta method can be used to obtain confidence limits for RE (Henson, 1975), with the former typically being preferred.

Burzykowski, Molenberghs and Buyse (2005) provided $(1 - \alpha)\%$ confidence limits for *PTE* using Fieller's theorem:

$$1 - \frac{B \pm \sqrt{B^2 - AC}}{A}$$

where

$$A = \hat{\beta}^2 - z_{\alpha/2}^2 \text{Var}(\hat{\beta})$$

$$B = \hat{\beta}\hat{\beta}_\gamma - z_{\alpha/2}^2 \text{Cov}(\hat{\beta}\hat{\beta}_\gamma)$$

and

$$C = \hat{\beta}_\gamma^2 - z_{\alpha/2}^2 \text{Var}(\hat{\beta}_\gamma)$$

and $z_{\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the normal distribution. Similarly, the $(1 - \alpha)\%$ confidence limits for *RE* is:

$$\frac{B \pm \sqrt{B^2 - AC}}{A}$$

where

$$A = \hat{\alpha}^2 - z_{\alpha/2}^2 \text{Var}(\hat{\alpha})$$

$$B = \hat{\alpha}\hat{\beta} - z_{\alpha/2}^2 \text{Cov}(\hat{\alpha}, \hat{\beta})$$

and

$$C = \hat{\beta}^2 - z_{\alpha/2}^2 \text{Var}(\hat{\beta})$$

1.3 Advantages of Bayesian Approaches in Surrogate Endpoints Evaluation

A Bayesian analysis of surrogate endpoints avoids the potential difficulties associated with frequentist hypothesis testing and maximum likelihood estimation. For example, in the frequentist approach, one initially tests for interaction between *T* and *S*.

Interaction is assumed to be zero if one fails to reject the hypothesis of no interaction. We shall see in Chapter 4 that this is problematic in the presence of small but practically significant interaction. As another example, maximum likelihood estimation routinely produces estimates of *PTE* that are outside the unit interval.

Producing a posterior distribution on *PTE*, *RE*, or ρ , conditional on available data, enables us straight-forward conclusions such as “The posterior probability that *PTE* > 0.75 is 0.85”. Cowles (2002) summarized three advantages of Bayesian approaches to surrogate endpoint validation:

1. Estimating variances of parameters needed by Fieller’s method or the delta method can be very difficult, especially with small sample sizes;
2. The Bayesian approach allows improvement of precision in estimation by introducing prior information;
3. Complex models are easily fit using Markov chain Monte Carlo (MCMC) methods.

CHAPTER TWO

Bayesian Evaluation of Surrogate Endpoints Using *PTE*

As discussed in Chapter 1, to reduce time, size and cost of clinical trials, surrogate endpoints are often measured instead of true endpoints. Surrogate endpoints are intermediate biomarkers that can be measured easier and earlier. However, an invalid surrogate endpoint may not provide reliable evidence about effects of the intervention on true endpoints. Freedman, Graubard, and Schatzkin (1992) suggested focusing on the proportion of the treatment effect explained by surrogate endpoints (*PTE*) as one validation criterion for surrogate endpoints. A good surrogate endpoint should be able to explain a large portion of the treatment effect.

This chapter considers criteria for validating surrogate endpoints based on *PTE*. Section 2.1 details the frequentist and Bayesian inferences for *PTE*. Advantages and disadvantages of *PTE* are also discussed in this section. We propose the conditional prior approach to improve *PTE* estimation in Section 2.2. In Section 2.3, simulations are conducted to compare our conditional prior approach with frequentist methods and a Bayesian method proposed by Cowels (2002). We investigate the induced prior structure for *PTE* and its influence on the posterior distribution in Section 2.3.

2.1. Introduction to PTE

In this section, we introduce both frequentist and Bayesian methods for estimating *PTE*. Weaknesses inherent in current *PTE* estimation methods are also detailed in this section.

2.1.1 Inferences for PTE

PTE is intuitively attractive due to its simplicity. Denote by E the true biological endpoint, such as HIV infection or a bone fracture. Let T be the treatment effect and let S be a surrogate endpoint, such as CD4 counts or bone mineral density measurements. Consider the models

$$\mathcal{E}(E|T, S) = \mu_\gamma + \beta_\gamma T + \gamma S \quad (2.1)$$

and

$$\mathcal{E}(E|T) = \mu + \beta T \quad (2.2)$$

The proportion of the treatment effect explained by a surrogate endpoint is defined as:

$$PTE = 1 - \frac{\beta_\gamma}{\beta}.$$

Freedman, *et al.* (1992) proposed a maximum likelihood estimator for *PTE*. Cowles (2002) introduced a Bayesian method to estimate *PTE* in a general linear model framework. We consider both in detail in subsequent sections.

2.1.2 Maximum Likelihood Estimation of PTE

PTE is the most widely used criterion due to its simplicity and ease of interpretation. The closer *PTE* is to 1, the better the surrogate endpoint. Frequentist methods use a maximum likelihood estimator (MLE) of *PTE*. Since *PTE* is a ratio of two parameters, its confidence limits can be obtained either by Fieller's theorem or by the delta method. Herson (1975) provided details on the use of Fieller's theorem and the delta method to calculate significance intervals for such ratios.

Freedman *et al.* (1992) proposed the first statistical method to estimate *PTE* by modeling two logistic regression equations simultaneously:

$$\text{Logit}[P(E = 1/T = j)] = \mu + \tau_j$$

and

$$\text{Logit}[P(E = 1/S = s_i; T = j)] = \mu + \sigma_i + \tau_{ja}$$

where E is the binary response, T is the binary treatment indicator, μ is the overall mean, τ_j is the effect of the j^{th} treatment ($j = 0, 1$), σ_i is the effect of the i th marker level ($i = 1, 2, \dots, k$), and τ_{ja} is the effect of the j^{th} treatment adjusted for the effect of the marker. The authors estimated PTE as

$$\widehat{PTE}_{FGS} = 1 - \frac{\hat{\tau}_{1a}}{\hat{\tau}_1}$$

where $\hat{\tau}_{1a}$ and $\hat{\tau}_1$ are MLE's for τ_{1a} and τ_1 , respectively. They suggested that a surrogate be considered valid if the lower 95% confidence limit exceeds a pre-chosen critical value, say 0.75.

Lin, Fleming and DeGruttola (1997) estimated PTE in another context. A Cox proportional hazards model was used to model the time to occurrence of a clinical endpoint. If T is the treatment indicator variable and $S(t)$ is a vector of covariates representing the history of the surrogate marker, the two models of interest are

$$\lambda(E | T) = \lambda_{10}(t)e^{\beta_1 T}$$

and

$$\lambda(E | T, S) = \lambda_{20}(t)e^{\beta_{1a} T + \gamma S(t)},$$

where $\lambda_{10}(t)$ and $\lambda_{20}(t)$ are the baseline hazard functions. Under this model PTE may be estimated as

$$\widehat{PTE}_{LFD} = 1 - \frac{\hat{\beta}_{1a}}{\hat{\beta}_1}$$

Freedman, *et al.* (1992) and Lin, *et al.* (1997) used either the delta method or Fieller's theorem to compute their confidence intervals for PTE . This results in

considerably wide intervals unless the unadjusted treatment effect is highly significant. One way to define “highly significant” in this context is to require that $\beta_1/\text{SE}(\beta_1)$ be greater than 4, which is a very strong condition.

2.1.3 Cowles’ Bayesian Method for Estimation of PTE

Cowles (2002) proposed estimation of *PTE* in the context of a generalized linear model. Let E_i be the clinical outcome of the i^{th} patient, $i = 1, \dots, n$. We assume E_1, \dots, E_n are independently distributed from an exponential family with density function of the form

$$f(E_i; \theta_i, \phi) = \exp \left\{ \frac{E_i \theta_i - b(\theta_i)}{a(\phi)} + c(E_i, \phi) \right\}$$

Let \mathbf{z}_i be the vector of covariates for subject i , $\boldsymbol{\beta}$ be a vector of coefficients and g be the “canonical link” which is a monotonic and differentiable function. We can rewrite the expectation of E_i as a function of \mathbf{z}_i :

$$\mathcal{E}(E_i) = g^{-1}(\mathbf{z}_i^T \boldsymbol{\beta}).$$

To compute *PTE*, the “full model” is defined to be the model which includes the surrogate marker:

$$\mathcal{E}(E_i | S_i, T_i) = g^{-1}(\mu_\gamma + \beta_\gamma T_i + \gamma S_i) \quad (2.3)$$

where T_i is the treatment indicator and S_i is the marker value. The “reduced” model omits the marker term from equation (2.3):

$$\mathcal{E}(E_i | T_i) = g^{-1}(\mu + \beta T_i). \quad (2.4)$$

If both the full and reduced models hold, then *PTE* is estimated as

$$\widehat{PTE} = 1 - \frac{\hat{\beta}_\gamma}{\hat{\beta}}. \quad (2.5)$$

Furthermore, if both the full model and the reduced model hold, they are related implicitly as

$$g^{-1}(\mu + \beta T_i) = \int g^{-1}(\mu_\gamma + \beta_\gamma T_i + \gamma S) f(S | T_i, \mu_S, \beta_S) dS, \quad (2.6)$$

where function $f(\cdot)$ models the association between surrogate endpoint and treatment effect and has the following relation:

$$\mathcal{E}(S|T) = f^{-1}(\mu_S + \beta_S T_i).$$

The joint posterior distribution for $(\mu_\gamma, \beta_\gamma, \gamma, \beta)$ can be divided into two factors: A marginal joint posterior distribution for parameters in the full model and a conditional posterior distribution based on the full model using (2.6),

$$p(\mu_\gamma, \beta_\gamma, \gamma, \beta | E, T, S) = p(\mu_\gamma, \beta_\gamma, \gamma | E, T, S) p(\beta | \mu_\gamma, \beta_\gamma, \gamma, E, T, S). \quad (2.7)$$

Typically, this posterior will not be of closed form. Markov chain Monte Carlo (MCMC) methods enable samples to be drawn from otherwise intractable posterior distributions in Bayesian modeling. At convergence, the MCMC simulation produces a sequence of values from the joint posterior. Pairs of values for (β, β_γ) from this sequence can then be used to construct the posterior distribution of *PTE*.

2.1.4 Disadvantages of *PTE*

For a small clinical trial with rare number of clinical events, the interval estimates for *PTE* will be wide. That is, *PTE* is highly variable. Large random trials or a meta-analysis of many related trials can improve interval estimation for *PTE*. Burzykowski, Molenberghs and Buyse (2005) pointed out that, even with large numbers of observations, the denominator of *PTE* (the effect of treatment upon the true endpoint) will be estimated with little precision. Otherwise, a surrogate endpoint would not be

needed. In addition, estimates of *PTE* might not be in the unit interval if the direction of the treatment effect, T , on the true endpoint, E , is reversed after adjustment for the surrogate endpoint, S . In this instance, it will be difficult to interpret *PTE* since it is supposed to be a proportion.

Despite these shortcomings, *PTE* is still widely used, principally due to its simplicity. Maximum likelihood estimation (as presented above) remains the most widely employed method for providing inferences about *PTE*. Researchers have proposed several non-Bayesian methods to improve estimation of *PTE* including Freedman (2001), Li Z, Meredith and Hoseyni (2001), Wang and Taylor (2002), and Chen, Wang and Snapinn (2003).

The Bayesian approach developed by Cowles (2002) also presents difficulties. As noted above, she mentioned ruling out interaction before proceeding with her Bayesian analysis. However, she did not suggest a Bayesian method for doing so. Also, utilizing diffuse priors can produce posterior estimates of *PTE* that are not in the unit interval. In the next section, we introduce a new Bayesian approach based on a conditional prior structure. Our approach is free of the problems outlined above.

2.2. Conditional Prior Approach

Cowels (2002) introduced two prior structures for the parameters in the model defined by (2.3) and (2.4)—locally uniform priors and informative normal priors, the latter based on historical data. In her application, a Cox proportional hazards model (a special case of equation (2.3)) was built. For comparison, she calculated a frequentist estimator for *PTE* using the method in Lin *et al.* (1997).

In her Bayesian approach for estimating *PTE*, she first assumed a locally uniform prior structure on all model coefficients. She also used an informative multivariate normal-Wishart prior on the parameters from the full model based on historical data.

Cowels noted that posterior estimates of *PTE* might be outside the unit interval when using non-informative priors. The use of non-informative priors also yields unacceptably wide interval estimates of *PTE*. Motivated by these problems we have developed the conditional prior structure introduced in the next section.

2.2.1 Assumptions

There are two assumptions underlying our conditional prior method.

1. After adjusting for the effect of a surrogate endpoint, the treatment's effect on the true endpoint is no larger than the overall treatment effect. Theoretically, the overall treatment effect can be separated into two parts: the effect explained by the surrogate and the effect attributed to other factors associated with the treatment. It is therefore reasonable to assume, in models (2.1) and (2.2), that $|\beta_\gamma| \leq |\beta|$.
2. Adjusting for the surrogate will not change the direction of the treatment effect on the true endpoint. That is, $0 \leq \beta_\gamma \leq \beta$ or $\beta \leq \beta_\gamma \leq 0$.

In the conditional prior structure we detail below, let $p(\beta)$ be the prior distribution for β . We construct a prior, $p(\beta_\gamma | \beta)$, conditional on β .

2.2.2 Conditional Prior Structure for β and β_γ

We considered two families of distributions as candidates for our conditional prior structure. Since Cowels (2002) used a normal prior structure for β and β_γ , we

considered truncated normal distributions in one approach. In our second approach, we used beta distributions because of both their flexibility in shape and their finite support.

2.3 A Small Sample Simulation Study

To investigate the efficacy of using either truncated normal distributions or beta distributions for our conditional prior structure, we conducted a small simulation study involving a small sample size of only 40 patients. We chose this small size in this preliminary study to highlight certain problems with the maximum likelihood approach. Larger, more realistic sample sizes are considered in Chapter 3.

2.3.1 Data Generation

We begin by comparing our conditional priors, based on either truncated normal or beta distributions, to the use of diffuse normal priors in Cowles (2002). To this end a small simulation study was conducted. Forty patients, half in the treatment group and half in the control group, were simulated 100 times. Values of the surrogate endpoint, S , were generated from a normal distribution with mean 1 for the treatment group and mean -1 for control group. Both were given a variance of 1. The treatment indicator variable, T , equals to -0.5 for control group and 0.5 for treatment group. The true endpoint, E , was simulated from a Poisson distribution with mean $\exp(\beta_\gamma T + \gamma S)$. Hence, the likelihood function, l_1 , for model (2.1) is:

$$l_1(\mu_\gamma, \beta_\gamma, \gamma | \mathbf{E}, \mathbf{T}, \mathbf{S}) = \sum_{i=1}^n [-\mu_{1i} + E_i \log(\mu_{1i}) - \log(E_i)], \quad (2.8)$$

where $\mathbf{E} \equiv (E_1, \dots, E_n)$ is the vector of true endpoint values, $\mathbf{T} \equiv (T_1, \dots, T_n)$ is the vector of treatment indicator values, $\mathbf{S} \equiv (S_1, \dots, S_n)$ is the vector of surrogate endpoint values, and

$$\mu_{1i} = \mu_\gamma + \beta_\gamma T_i + \gamma S_i.$$

The likelihood function, l_2 , for (2.2) is:

$$l_2(\mu, \beta | \mathbf{E}, \mathbf{T}) = \sum_{i=1}^n [-\mu_{2i} + \mathbf{E}_i \log(\mu_{2i}) - \log(\mathbf{E}_i)], \quad (2.9)$$

where $\mu_{2i} = \mu + \beta T_i$. In our simulation study, we used SAS to generate data using $\beta_\gamma = 0.75$ and $\gamma = 0.375$, yielding $PTE = 0.5$. The SAS program is included in Appendix B.1.

With our conditional prior structure, the posterior distributions have the form

$$p(\mu, \beta | \mathbf{E}, \mathbf{T}) \propto l_1(\mu, \beta | \mathbf{E}, \mathbf{T}) p(\beta) p(\mu) \quad (2.10)$$

and

$$p(\mu_\gamma, \beta_\gamma, \gamma | \mathbf{E}, \mathbf{T}, \mathbf{S}) \propto l_2(\mu_\gamma, \beta_\gamma, \gamma | \mathbf{E}, \mathbf{T}, \mathbf{S}) p(\beta_\gamma | \beta) p(\mu_\gamma) \quad (2.11)$$

For our applications these posterior distributions were not of closed form. This would typically be the case in practice. For this reason we used MCMC methods, implemented with the software package WinBUGS (Spiegelhalter, Thomas, Best, and Lunn 2003). To facilitate writing simulations requiring MCMC methods, the R2WinBUGS package provides convenient functions to call WinBUGS from R. It automatically writes the data and scripts in a format readable by WinBUGS for processing in batch mode. After the WinBUGS process has finished, it is possible either to read the resulting data into R or to use the facilities of the coda package for further analyses of the output (Sturtz, Ligges and Gelman, 2005).

2.3.2 Prior Structure for β and β_γ

As in Cowles' (2002) model, we take μ , μ_γ , and γ to have independent, diffuse normal priors, specifically, $N(0, 100,000)$. We studied three prior structures for β and β_γ in the simulation study.

Let $N_{[\nu, \varpi]}(\mu, \sigma^2)$ be a normal distribution with mean μ and variance σ^2 that is truncated and scaled to have finite support $[\nu, \varpi]$, $\nu < \varpi$. This distribution has density

$$f(x | \mu, \sigma^2, u, v) = \frac{h}{\sqrt{2\pi\sigma^2}} \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^2\right\} I_{[u,v]}(x)$$

where $I_{[\nu, \varpi]}(\xi)$ is the indicator function and

$$h^{-1} = \int_u^v \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx.$$

Lunn (2003) developed an add-in component for WinBUGS to sample from the truncated normal distribution.

Denote by $\text{beta}_{[u,v]}(a, b)$ a beta density translated and scaled to have support $[u, v]$, $u < v$. This distribution has density

$$f(x | a, b, u, v) = \frac{(v-u)^{1-a-b}}{B(a, b)} (x-u)^{a-1} (v-x)^{b-1}$$

where

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt.$$

The three prior structures for β and β_γ used in our simulation are as follows:

- (1) Diffuse normal structure: $\beta \sim N(0, 100,000)$, $\beta_\gamma \sim N(0, 100,000)$
- (2) Truncated Normal structure: $\beta \sim N_{[0, 1000]}(0, 100,000)$, $\beta_\gamma \sim N_{[0, \beta]}(0, 100,000)$

(3) Generalized beta structure: $\beta \sim \text{beta}_{[0, d]}(a, b)$, $\beta_\gamma \sim \text{beta}_{[0, \beta]}(f, g)$,

where a, b, d, f , and g , are to be specified.

The truncated diffuse normal prior structure in (1) is close to a uniform distribution on the interval $[0, 1,000]$. In our study, we used $a = b = f = g = 3$ and $d = 5$. The motivations for choosing such values will be discussed in the next section.

Each of the three prior structures detailed above induces a prior on PTE . In Section 2.4 we shall identify the exact distribution of the induced prior for PTE under prior structures (1) and (3). (It will turn out that prior structure (2) is not very useful.)

We used the SAS GENMOD procedure to fit the Poisson regression of $E|T$ and $E|T, S$ simultaneously. The maximum likelihood estimators (MLE's) for β and β_γ were estimated. The MLE for PTE is the ratio of MLE's for β_γ and β in model (2.1) and (2.2). The SAS program is included in Appendix C.1.

The posterior distributions, (2.10) and (2.11), under prior structure (3) appears to be intractable. As noted earlier, we used the WinBUGS package to implement MCMC methods for obtaining the posterior distributions. R Codes for these models are included in Appendix D.1, D.2 and D.3. WinBUGS model is included in Appendix E.1.

For each of the 100 samples, we based our conclusions on 10,000 iterations each from two chains beginning with dispersed starting values and a burn in of 1000 iterations per chain. For convergence diagnostic purposes, we computed Gelman-Rubin statistics and examined autocorrelation plots. We found no problems with convergence.

2.3.3 Simulation Results

Posterior kernel density plots typical of those obtained in this simulation study are exhibited in Figure 2. In Figure 2(a), note that the posterior for PTE using diffuse

normal priors has support outside of $[0, 1]$. The truncated normal and beta conditional priors do not suffer from this problem, as can be seen in Figures 2(b) and 2(c).

In this simulation, $PTE = 0.5$. The posterior modal values in all three plots are seen to be less than this true value. Understanding this discrepancy depends on understanding the role prior information has played in the posterior. We shall address this issue in the next section.

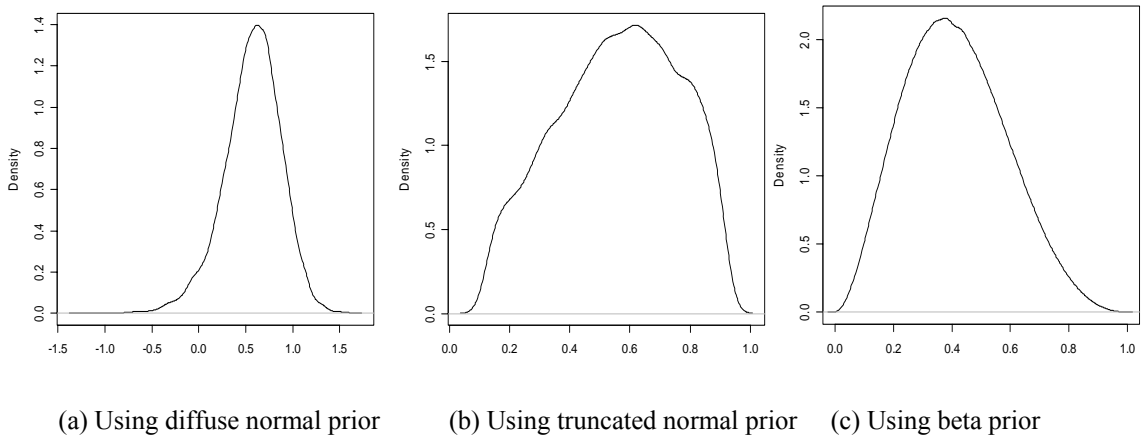


Figure 2. Posterior distribution of PTE using diffuse normal prior structure.

In Table 1 we compare the four methods based on 100 simulated data sets. The first row shows the mean and median of the 100 MLE estimators. Of the 100 MLE's computed, 45 were outside of the unit interval. While this is in part due to the small sample size, we shall see in the next chapter that, even with larger sample sizes, the MLE frequently produces estimates outside the unit interval. The remaining rows indicate the mean and median of the 100 posterior medians for the PTE using three different prior structures as well as the 95% percentiles for the 100 posterior median PTE 's. Analysis with the diffuse normal also produced estimates of PTE outside $[0,1]$, with 49 out of 100

simulations doing so. Again, this is partly due to the small sample size and, again, we shall see that even larger sample sizes can yield such results.

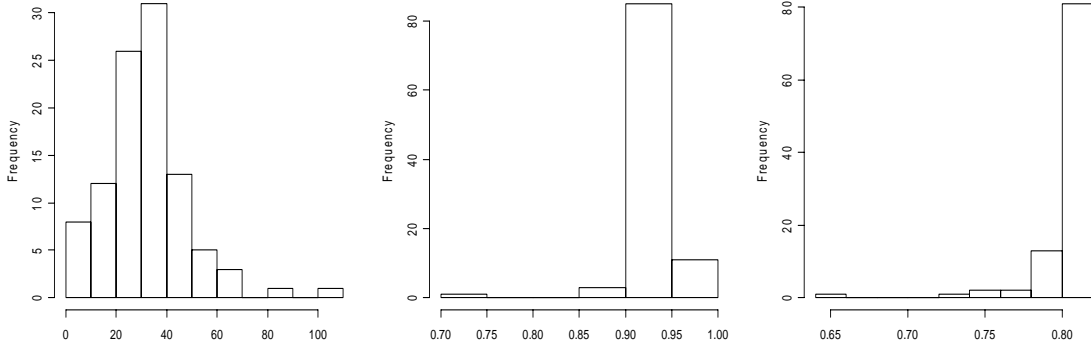
For the truncated normal prior structure, β_γ is restricted to be no larger than β and both parameters are assumed to be positive. Under the truncated normal and the conditional beta prior structure, the posterior for *PTE* has unit interval support. The posterior means and medians for *PTE* were close to the true value than the MLE or the means and median under either diffuse normal.

Table 1. Comparison of simulation results. True *PTE* = 0.5. Entries are the mean and percentiles of the posterior median and mean for the 100 samples.

Interval Type		mean	2.5%	median	97.5%
MLE		0.06	−0.02	0.051	0.11
Diffuse normal	median	0.57	0.70	0.60	2.06
	mean	−3.47	−22.3	0.21	11.08
Conditional truncated normal	median	0.50	0.33	0.53	0.58
	mean	0.50	0.37	0.52	0.55
Conditional beta	median	0.51	0.40	0.528	0.57
	mean	0.51	0.41	0.524	0.55

The widths of 95% credible sets for *PTE* for the 100 samples are exhibited in Figure 3. Not surprisingly, the more informative conditional beta priors yield relatively tighter intervals, but intervals for all methods are quite wide, as is typical for *PTE*. With larger sample sizes, the interval widths are narrower, but not satisfactorily so, as we shall see. (Confidence interval widths are also typically quite wide but are not depicted here.)

Of course, better results may be obtained by increasing prior precision or the size of the trial. One may also combine several studies in a meta-analysis. We consider larger simulated samples in Section 3.2. Additionally, we look at a method for using historical data in fitting informative priors in Section 3.3.



(a) Using diffuse normal prior (b) Using truncated normal prior (c) Using conditional Beta prior
Figure 3. Histograms of the confidence interval widths from the 100 simulated samples.

It appears that there is little to be gained by using conditional truncated normal priors over conditional beta priors. We chose to focus on conditional beta priors in the rest of the dissertation.

2.4 Induced Prior Structure for PTE

In the simulation study, we placed a prior structure on β and β_γ thereby inducing a prior on PTE . Knowing the induced prior for PTE will allow us to gauge prior-to-posterior changes. Furthermore, distinguishing the relationship between prior for β and β_γ and prior for PTE can help us to elicit the prior parameters.

In this section, we consider the induced prior for PTE based on two prior structures for β and β_γ : conditional beta priors and diffuse normal priors.

2.4.1 Induced Prior for PTE—Ratio of Two Dependent Beta Distributions ($\beta \geq \beta_\gamma \geq 0$)

The conditional beta prior structure, $\beta \sim \text{beta}_{[c, d]}(a, b)$ and $\beta_\gamma | \beta \sim \text{beta}_{[e, \beta]}(f, g)$, yields the prior densities

$$f(\beta) = \frac{(d-c)^{1-a-d} (\beta-c)^{a-1} (d-\beta)^{b-1}}{B(a, b)}, \quad (2.12)$$

where $c < \beta < d$, and

$$f(\beta_\gamma | \beta) = \frac{(\beta - e)^{1-f-g} (\beta_\gamma - e)^{f-1} (\beta - \beta_\gamma)^{g-1}}{B(f, g)}, \quad (2.13)$$

where $a < e < \beta_\gamma < \beta$.

From (2.12) and (2.13), we can derive the joint distribution for β and β_γ :

$$\begin{aligned} f(\beta_\gamma, \beta) &= \frac{(\beta - e)^{1-f-g} (\beta_\gamma - e)^{f-1} (\beta - \beta_\gamma)^{g-1}}{B(f, g)} \\ &\quad \times \frac{(d - c)^{1-a-b} (\beta - c)^{a-1} (d - \beta)^{b-1}}{B(a, b)}, \end{aligned} \quad (2.14)$$

where $c < e < \beta_\gamma < \beta$ and $c < \beta < d$.

We now derive the induced prior for *PTE*. For simplicity in these derivations, let $P \equiv PTE = 1 - \beta_\gamma/\beta$. Let $W = \beta$ so that $\beta_\gamma = (1 - P)W$. After applying the Jacobean of transformation, we obtain

$$\begin{aligned} f(P, W) &= \frac{|W| (W - e)^{1-f-g} ((1 - P)W - e)^{f-1} (W - (1 - P)W)^{g-1}}{B(f, g)} \\ &\quad \times \frac{(d - c)^{1-a-b} (\beta - c)^{a-1} (d - \beta)^{b-1}}{B(a, b)}, \end{aligned} \quad (2.15)$$

where $0 < P < 1$ and $c < W < d$. The Jacobean matrix is

$$J = \begin{vmatrix} \frac{\partial \beta}{\partial W} & \frac{\partial \beta}{\partial P} \\ \frac{\partial \beta_\gamma}{\partial W} & \frac{\partial \beta_\gamma}{\partial P} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ (1 - P) & -W \end{vmatrix} = |W|.$$

After some simple algebra (2.15) becomes

$$f(P, W) = \frac{1}{B(f, g)B(a, b)} (d - c)^{1-a-b} P^{g-1} (1 - P)^{f-1} |W| (W - e)^{1-f-g} \times \left(W - \frac{e}{1-P} \right)^{f-1} W^{g-1} (W - c)^{a-1} (d - W)^{b-1}. \quad (2.16)$$

There are two scenarios in calculating the induced prior for *PTE*:

Scenario 1: $c = e = 0$, which means $\beta \sim \text{beta}_{[0, d]}(a, b)$ and $\beta_\gamma | \beta \sim \text{beta}_{[0, \beta]}(f, g)$ and

$$f(P, W) = \frac{P^{g-1} (1 - P)^{f-1} (d)^{1-a-b} (W)^{a-1} (d - W)^{b-1}}{B(a, b)B(f, g)}, \quad (2.17)$$

By integrating out W in (2.17), it is easy to show that P follows a standard beta distribution with shape parameter g and f , that is,

$$\beta \sim \text{beta}_{[0, d]}(a, b) \text{ and } \beta_\gamma | \beta \sim \text{beta}_{[0, \beta]}(f, g) \Rightarrow 1 - \frac{\beta}{\beta_\gamma} \sim \text{beta}_{[0, 1]}(g, f).$$

It may seem that the prior for *PTE* is independent of the prior for β . In fact, the shape parameters f and g used in the prior for β_γ are highly dependent on the prior structure for β . If we have the prior information for β , the induced prior for *PTE* depends on β through β_γ since

$$E(\beta_\gamma | \beta) = \beta \frac{f}{f + g}$$

and

$$\text{Var}(\beta_\gamma | \beta) = \beta^2 \frac{fg}{(f + g + 1)(f + g)}.$$

Methods for the elicitation of shape parameters for beta distributions have been considered, for example, by Chaloner and Duncan (1983). Choices for scale parameters are discussed in the next section.

Scenario 2: When $c \neq 0$ and $e \neq 0$, the induced prior for PTE must be determined by simulation methods.

2.4.2 Induced Prior for PTE —Ratio of Two Dependent Beta Distributions ($\beta \leq \beta_\gamma \leq 0$)

In the case of $\beta \leq \beta_\gamma \leq 0$, $\beta \sim \text{beta}_{[c, d]}(a, b)$ and $\beta_\gamma | \beta \sim \text{beta}_{[\beta, e]}(f, g)$, their density functions are:

$$f(\beta) = \frac{(d-c)^{1-a-b} (\beta-c)^{a-1} (d-\beta)^{b-1}}{B(a, b)},$$

where $c < \beta < d$ and

$$f(\beta_\gamma | \beta) = \frac{(e-\beta)^{1-f-g} (\beta_\gamma - \beta)^{f-1} (e-\beta_\gamma)^{g-1}}{B(f, g)},$$

where $\beta < \beta_\gamma < e$. We can derive the joint distribution for β and β_γ :

$$f(\beta_\gamma, \beta) = \frac{(e-\beta)^{1-f-g} (\beta_\gamma - \beta)^{f-1} (e-\beta_\gamma)^{g-1}}{B(f, g)} \times \frac{(d-c)^{1-a-b} (\beta-c)^{a-1} (d-\beta)^{b-1}}{B(a, b)},$$

where $\beta < \beta_\gamma < e$ and $c < \beta < d$.

Again, let $P \equiv PTE = 1 - \beta_\gamma/\beta$ and $W = \beta$ so that $\beta_\gamma = (1-P)W$. We have:

$$f(P, W) = \frac{|W| (e-W)^{1-f-g} [(1-P)W - W]^{f-1} [e - (1-P)W]^{g-1}}{B(f, g)} \times \frac{(d-c)^{1-a-b} (W-c)^{a-1} (d-W)^{b-1}}{B(a, b)},$$

where $0 < P < 1$ and $c < W < d$. The Jacobean matrix is

$$J = \begin{vmatrix} \frac{\partial \beta}{\partial W} & \frac{\partial \beta}{\partial P} \\ \frac{\partial \beta_\gamma}{\partial W} & \frac{\partial \beta_\gamma}{\partial P} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ (1-P) & -W \end{vmatrix} = |W|.$$

When $d = e = 0$, which means $\beta \sim \text{beta}_{[c, 0]}(a, b)$ and $\beta_\gamma | \beta \sim \text{beta}_{[\beta, 0]}(f, g)$

$$f(P, W) = \frac{P^{f-1} (1-P)^{g-1} (-W)^{b-1} (W-c)^{a-1} (-c)^{1-a-b}}{B(a, b) B(f, g)}, \quad (2.18)$$

Integrating out W in (2.18) we obtain $P \sim \text{beta}_{[0,1]}(f, g)$. Hence, the induced prior distribution for PTE follows a standard beta with shape parameter f and g . That is

$$\beta \sim \text{beta}_{[c, 0]}(a, b) \text{ and } \beta_\gamma | \beta \sim \text{beta}_{[\beta, 0]}(f, g) \Rightarrow 1 - \frac{\beta}{\beta_\gamma} \sim \text{beta}_{[0,1]}(f, g).$$

2.4.3 Example Using Conditional Beta Priors and the Induced Prior on PTE

For this example we simulated a data set with 200 patients. Half of the patients were allocated to be in the treatment group and half in control group. Values for the surrogate endpoint, S , were generated from a normal distribution with mean 1 and variance 1 for the treatment group. Values of S for the control group were assumed to follow a normal distribution with mean -1 and variance 1. The treatment indicator variable, T , is set to -0.5 for the control group and 0.5 for the treatment group. The true endpoint, E , is simulated from a Poisson distribution with mean $\exp(\beta_\gamma T + \gamma S)$, where $\beta_\gamma = 0.75$ and $\gamma = 0.375$, yielding $PTE = 0.5$. All coefficients other than β and β_γ are assumed to follow the diffuse normal prior $N(0, 100,000)$.

In this example we used the conditional priors

$$\beta \sim \text{beta}_{[0, 5]}(5, 6)$$

and

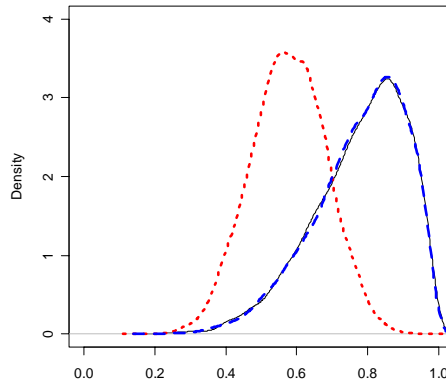
$$\beta_\gamma | \beta \sim \text{beta}_{[0, \beta]}(2, 7).$$

It follows that

$$PTE \sim \text{beta}_{[0,1]}(7, 2).$$

This is a relatively optimistic prior structure for PTE with a mean of 0.78. R-code that was used to simulate induced prior for PTE is attached as Appendix D.6.

The prior and posterior distributions for PTE are graphed in Figure 4. In our example, the MLE is 0.41 and the mode of the posterior is about 0.577. The mean and median of the posterior are 0.569 and 0.572, respectively. The mode, mean, and median of the posterior are quite close to the true value of 0.5, despite the overly optimistic prior given to PTE .



— simulated prior for PTE ---- derived induced prior for PTE posterior for PTE
Figure 4. Prior and posterior for PTE .

2.4.4 Induced Prior for PTE -- Ratio of Two Independent Normal Distributions

Suppose $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$ and X is independent of Y . Their joint distribution is

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp \left\{ - \left(\frac{x^2}{2\sigma_X^2} + \frac{y^2}{2\sigma_Y^2} \right) \right\}$$

The distribution of $U = X/Y$ is, of course, Cauchy with density

$$p(u) = \frac{1}{\pi} \frac{\sigma_X / \sigma_Y}{u^2 + (\sigma_X / \sigma_Y)^2}, -\infty < u < \infty. \quad (2.19)$$

In the case of $\sigma_X = \sigma_Y$, $p(u)$ becomes to be a standard Cauchy distribution, thereby placing a concentration of probability around zero.

In our example, $\beta \sim N(0, 100,000)$, $\beta_\gamma \sim N(0, 100,000)$ and the prior for PTE is Cauchy with location parameter 1, scale parameter 1, and density

$$p(PTE) = \frac{1}{\pi} \frac{1}{(1 - PTE)^2 + 1}, \quad -\infty < PTE < \infty. \quad (2.20)$$

The median and mode of this prior for PTE is 1 with undefined mean and variance. That is, using a diffuse normal prior structure for β and β_γ , the induced prior for PTE is centered at 1 and has support $(-\infty, \infty)$, as depicted in Figure 5. Hence, it is not surprising that the posterior estimates for PTE under this prior are often outside $[0, 1]$. This is hardly appropriate as a prior for PTE .

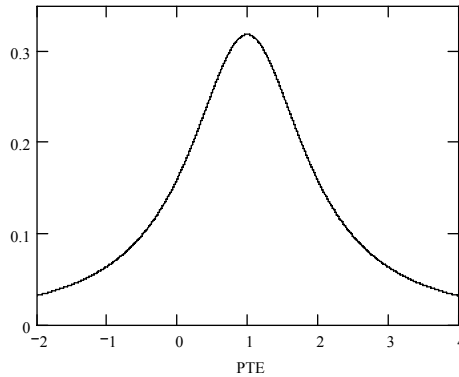


Figure 5. Induced Prior for PTE using diffuse normal prior for β and β_γ .

2.5 Support for Uniform Prior

The conditional prior structure uses $\beta \sim \text{beta}_{[0, d]}(a, b)$ and $\beta_\gamma | \beta \sim \text{beta}_{[0, \beta]}(f, g)$ in this section. We would like to investigate the relation between the support of the prior on β , determined by d , and the posterior of PTE . To this end we shall use the same data generated for use in Section 2.4.3. R-code is attached as Appendix D2 and WinBUGS Model is included as Appendix E1.

Figure 6 exhibits the relationship between features of the posterior and the upper bound, d , on the support of β in the conditional beta prior (3) introduced in Section 2.3.2. The posterior mean, median and 95% credible set for PTE are all shown in the plot. It is obvious that small values of d will profoundly influence the posterior of PTE . Use of such values should be based on very reliable prior knowledge. In general, it is recommended that d be given a relatively large value and the analyst should perform a prior-to-posterior sensitivity analysis. If both β and β_γ are negative, the impact of the lower bound c on posterior estimation needs to be studied in a similar fashion.

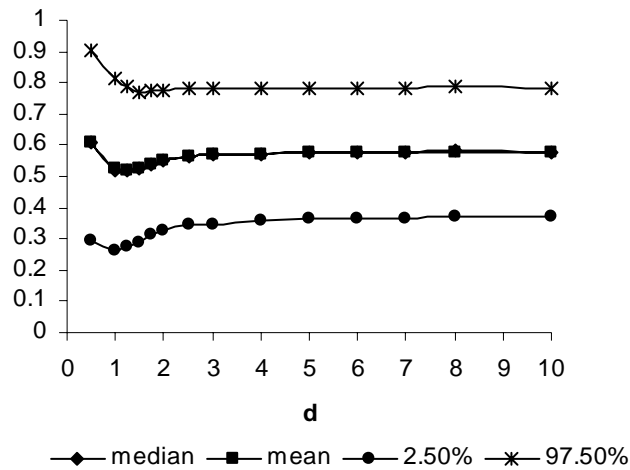


Figure 6. The posterior for PTE as a function of the bound, d , in $\beta \sim \text{beta}_{[0, d]}(a, b)$.

CHAPTER THREE

Simulation Study

In Chapter 2, we discussed Bayesian inference for *PTE* using conditional prior distributions. There we illustrated use of the method with a small simulation study, demonstrating, among other things, that the posterior for *PTE* using conditional priors will have the unit interval for support, unlike other existing methods.

After introducing the models used in this chapter's simulations in Section 3.1, we turn to the use of conditional priors in Section 3.2. Their performance is examined in a simulation using small ($n = 40$) and moderate ($n = 200$) sample sizes.

It was also seen in the last chapter that interval estimates of *PTE* are typically wide, regardless of the method used. This is a well-known problem with using *PTE* to assess surrogate markers. Burzykowski, Molenberghs and Buyes (2005) pointed out that there is substantial uncertainty about *PTE* if the treatment effect upon the true endpoint is small, especially for small clinical trial. Increasing the size of a study is one obvious way to counter this problem. Another is to combine studies in a meta-analysis. In Section 3.3, we present simulated examples in which informative priors are constructed from historical data, thereby narrowing subsequent posterior credible sets on *PTE*.

In Section 3.4 we consider prior-to-posterior sensitivity analyses. This is very important given the large data model variance component typically encountered in practice. Related MCMC convergence issues appear in Appendix A.

3.1 Introduction

As in Chapter 2, let E denote the true endpoint, T represent the treatment indicator, and let S be the surrogate endpoint. We relate E , T , and S with the models

$$\mathcal{E}(E|T, S) = \mu_\gamma + \beta_\gamma T + \gamma S \quad (3.1)$$

and

$$\mathcal{E}(E|T) = \mu + \beta T. \quad (3.2)$$

In our simulated data sets, we generate values of E using a Poisson distribution with mean

$$\lambda = \exp\{\beta_\gamma T + \gamma S\},$$

for various values of β_γ and γ . As before, the binary treatment indicator is defined as

$$T = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control}, \end{cases}$$

and we suppose S follows a normal distribution, conditional on the value of T :

$$S|T = -0.5 \sim N(-0.5, 0.5)$$

and

$$S|T = 0.5 \sim N(0.5, 0.5).$$

Poisson regression models were fit for E vs. T and S and E vs. T . We used SAS PROC GENMOD to fit the Poisson regression to obtain MLE's for β and β_γ , denoted by $\hat{\beta}$ and $\hat{\beta}_\gamma$, respectively. The SAS code we used is in Appendix C.1. Assuming that models (3.1) and (3.2) hold simultaneously, the MLE for PTE is then,

$$\widehat{PTE} = 1 - \frac{\hat{\beta}_\gamma}{\hat{\beta}}. \quad (3.3)$$

The Bayesian analyses in this chapter use MCMC methods as implemented in WinBUGS. R Code to call WinBUGS for these models is included in Appendix D.1 and WinBUGS model is included in Appendix E.1. The posterior distributions for β and β_γ are derived from the joint posterior distributions for $(\mu_\gamma, \beta_\gamma, \gamma)$ and (μ, β) ; that is, from

$$p_1(\mu_\gamma, \beta_\gamma, \gamma \mid \beta, \mathbf{E}, \mathbf{S}, \mathbf{T}) \propto l_1(\mu_\gamma, \beta_\gamma, \gamma \mid \beta, \mathbf{E}, \mathbf{S}, \mathbf{T}) p_0(\beta_\gamma \mid \beta) p_0(\mu_\gamma) p(\gamma) \quad (3.4)$$

and

$$p_2(\mu, \beta \mid \mathbf{E}, \mathbf{T}) \propto l_2(\mu, \beta \mid \mathbf{E}, \mathbf{T}) p_0(\mu) p_0(\beta) \quad (3.5)$$

where \mathbf{E} , \mathbf{T} , \mathbf{S} , l_1 , and l_2 are defined in (2.8) and (2.9).

We used the following prior structure on the coefficients in models (3.1) and (3.2). We placed independent diffuse normal distributions, $N(0, 100,000)$, on each of coefficients μ , μ_γ and γ . We used the following conditional priors on β and β_γ :

$$\beta \sim \text{beta}_{[0, d]}(a, b),$$

and

$$\beta_\gamma \mid \beta \sim \text{beta}_{[0, \beta]}(f, g)$$

In Section 2.4 we showed that the induced prior for *PTE* follows a $\text{beta}_{[0, 1]}(g, f)$. We chose $d = 5$ using a sensitivity analysis like that described in Section 2.5. The choices for a, b, f , and g are described below.

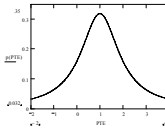
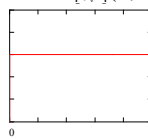
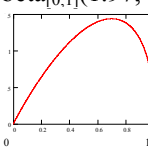
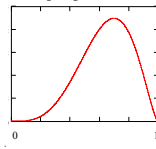
3.2 Simulation Study

3.2.1 Prior Structures

In this section, we will simulate data sets to study the performance of the conditional prior approach for small and moderate sample sizes. We also consider prior-

to-posterior robustness. The priors we consider are listed in Table 2, along with the corresponding induced priors for *PTE*. The latter are graphed in the table as well as in Figure 7(a) and (b).

Table 2. Prior structures for non-informative prior simulation.

Prior Structure			Inferences for <i>PTE</i>	
β and β_γ	β and β_γ	<i>PTE</i>	mean mode median	var std. dev.
Diffuse Normal	$\beta \sim N(0, 100,000)$ $\beta_\gamma \sim N(0, 100,000)$	$PTE \sim \text{Cauchy}$ 	undefined 1 1	undefined undefined
Uniform	$\beta \sim \text{beta}_{[0, 5]}(1, 1)$ $\beta_\gamma \sim \text{beta}_{[0, \beta]}(1, 1)$	$PTE \sim \text{beta}_{[0, 1]}(1, 1)$ 	0.5 0.5 0.5	$1/12 \approx 0.083$ 0.289
Beta Prior1	$\beta \sim \text{beta}_{[0, 5]}(2, 2)$ $\beta_\gamma \sim \text{beta}_{[0, \beta]}(1.42, 1.97)$	$PTE \sim \text{beta}_{[0, 1]}(1.97, 1.42)$ 	0.581 0.698 0.599	$1/18 \approx 0.056$ 0.235
Beta Prior2	$\beta \sim \text{beta}_{[0, 5]}(2, 2)$ $\beta_\gamma \sim \text{beta}_{[0, \beta]}(2.58, 4.68)$	$PTE \sim \text{beta}_{[0, 1]}(4.68, 2.58)$ 	0.645 0.7 0.659	$1/36 \approx 0.028$ 0.167

The diffuse normal and uniform priors are relatively non-informative, compared to the likelihood we ultimately employ. The two beta priors are informative. They were chosen so that the induced prior for *PTE* would have a mode of approximately 0.7 and a specified variance. This reflects some optimism about the validity of the surrogacy, which is attenuated by the choice of variance for these priors.

Beta prior 1 uses a variance of $1/18$. In that case, the 0.025 and 0.975 percentiles are 0.119 and 0.959, respectively, and the prior probability that PTE is at least 0.7 is approximately 0.356. For beta prior 2 the 0.025 and 0.975 percentiles are 0.294 and 0.921, respectively, and the prior probability that PTE is at least 0.7 is approximately 0.408.

We take $\beta \sim \text{beta}_{[0,5]}(2, 2)$. The upper bound on the support of this beta distribution, $d = 5$, is chosen using a sensitivity analysis as discussed in Section 2.5. The induced priors for PTE are shown in figure 7.

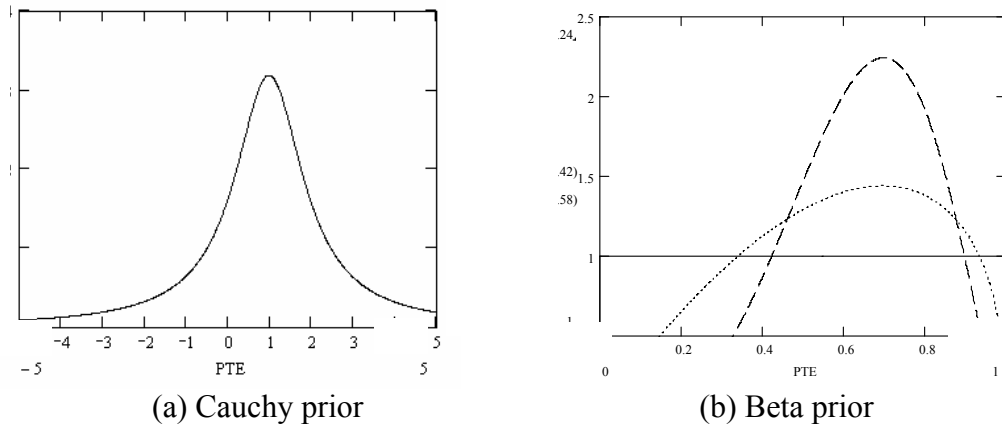


Figure 7. Induced priors for PTE .

3.2.2 Small Sample Size ($N = 40$)

To illustrate the performance of the conditional prior method with a small sample size, we simulated 40 patients, half in the treatment group and half in the control group. We generated 500 samples and used three “true” PTE values: 0.5, 0.7, and 0.9.

As in Chapter 2, WinBUGS was used to implement MCMC methods for computing the posteriors. In the first two scenarios ($PTE = 0.5$ and 0.7), two chains with dispersed initial values were used. Each chain began with 1,000 burn-in iterations followed by a sample of 10,000. No convergence problems were encountered. In the

third scenario ($PTE = 0.9$) we began with the same burn-in iterations and sample iterations, using two chains. However, on examining the autocorrelation plots, we detected problems with convergence of the chains. These were overcome by increasing the length of the chain and sampling from every 10th iteration after burn-in. In addition, we found it necessary in this scenario to reduce the variance in the independent normal priors to 10.

We now turn to the results of these simulations, considering each value of PTE .

$PTE = 0.5$ In this scenario, we generated 500 data sets each with 40 triples (E_i, T_i, S_i) using $\beta_\gamma = 0.75$ and $\gamma = 0.375$, yielding $PTE = 0.5$. Note that the informative priors in Table 2 (beta priors 1 and 2) have means, medians, and modes exceeding 0.5 and are therefore somewhat optimistic in this context. Table 3 displays averages of mean, median and 95% intervals of PTE estimations with four priors from 500 iterations respectively. The table 3 summarizes features of the 500 PTE posteriors. The numbers in the parenthesis indicate the 2.5% and 97.5% quantiles of the 500 estimators for PTE .

It is apparent that MLE's for PTE were frequently outside the unit interval and far from the true value $PTE = 0.5$. For the MLE's, the mean and median both tended to underestimate PTE . Fully 35% of the MLE's were outside the unit interval.

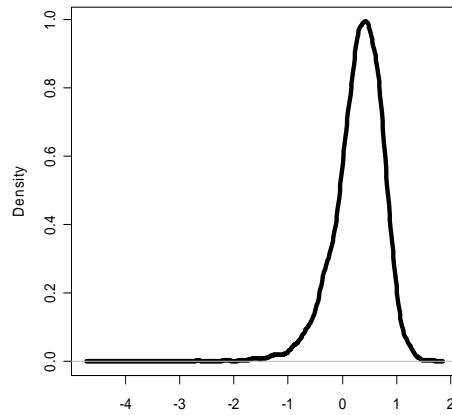
For the diffuse normal prior Bayesian analysis, 26.8% of the posterior means were outside the unit interval. Indeed, 95% of the lower credible set bounds were negative. Posterior means and medians underestimated PTE . Interval estimates were wide.

Bayesian analysis using conditional priors resulted in the posterior for PTE having support equal to the unit interval, by construction. Hence, no point or interval estimates were outside $[0, 1]$. Again, interval estimates were wide, as anticipated. Point

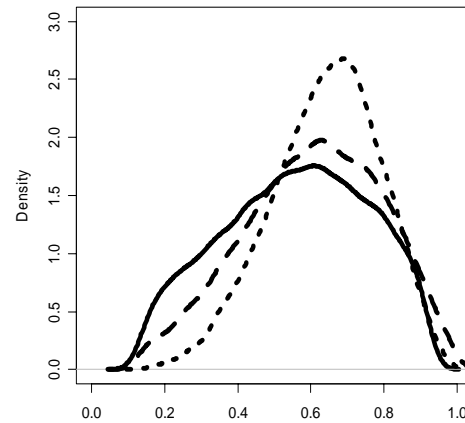
estimates were much closer to the true value of PTE , even for beta priors 1 and 2, which had the “optimistic” mode of 0.7.

Table 3. Simulation results with four prior structures when true $PTE = 0.5$ and $N = 40$. The average of the 500 MLE's was 0.219 with 2.5% and 97.5% quantiles -0.686 and 2.056 , respectively. The median of the MLE's was 0.453. Table entries summarize posterior features for the Bayesian analyses. Parenthetic entries are the 2.5% and 97.5% quantiles for 500 values.

Method	95% equal tail credible interval			
	mean	2.5%	median	97.5%
Diffuse	0.369	-1.131	0.471	1.329
Normal	$(-1.117, 2.440)$	$(-8.572, -0.113)$	$(-0.189, 1.396)$	$(0.691, 9.996)$
Uniform	0.502 $(0.315, 0.621)$	0.0298 $(0.0130, 0.0613)$	0.505 $(0.277, 0.670)$	0.967 $(0.797, 0.986)$
Beta Prior1	0.583 $(0.420, 0.677)$	0.136 $(0.077, 0.203)$	0.598 $(0.406, 0.716)$	0.950 $(0.831, 0.974)$
Beta Prior	0.639 $(0.536, 0.694)$	0.303 $(0.231, 0.354)$	0.651 $(0.535, 0.713)$	0.912 $(0.840, 0.935)$



(a) Using diffuse normal priors



(b) Using conditional priors

— using uniform prior for β and β_γ ; ---- using conditional beta prior 1; using conditional beta prior 2.

Figure 8. Posterior distributions for PTE .

Bayesian analysis using conditional priors resulted in the posterior for PTE having support equal to the unit interval, by construction. Hence, no point or interval

estimates were outside $[0, 1]$. Again, interval estimates were wide, as anticipated. Point estimates were much closer to the true value of PTE , even for beta priors 1 and 2, which had the “optimistic” mode of 0.7.

Figure 8 displays the posterior distributions for PTE with four prior structures. Results for other scenarios were similar.

$PTE = 0.7$ In this scenario, we generated 500 data sets each with 40 triples (E_i, T_i, S_i) using $\beta_\gamma = 0.375$ and $\gamma = 0.875$, yielding $PTE = 0.7$. Table 4 displays averages of mean, median and 95% intervals of PTE estimations with four priors from 500 iterations. The numbers in parentheses are the 95% percentiles for the corresponding estimator in 500 samples.

Table 4. Simulation results with four prior structures when true $PTE = 0.7$ and $N = 40$. The average of the 500 MLE's was 0.746 with 2.5% and 97.5% quantiles 0.197 and 1.597, respectively. The median of the MLE's was 0.691. Table entries summarize posterior features for the Bayesian analyses. Parenthetic entries are the 2.5% and 97.5% quantiles for 500 values.

Method	95% equal tail credible interval			
	mean	2.5%	median	97.5%
Diffuse	0.657	-0.291	0.688	1.394
Normal	(0.080, 1.757)	(-1.789, 0.497)	(0.188, 1.569)	(0.761, 4.618)
Uniform	0.578	0.0611	0.604	0.974
	(0.333, 0.732)	(0.018, 0.183)	(0.310, 0.790)	(0.7695, 0.992)
Beta	0.630	0.189	0.652	0.957
Prior 1	(0.432, 0.745)	(0.089, 0.311)	(0.425, 0.783)	(0.791, 0.980)
Beta	0.663	0.338	0.676	0.918
Prior 2	(0.530, 0.731)	(0.242, 0.412)	(0.529, 0.750)	(0.810, 0.945)

The 97.5% quantile of the MLE's for PTE exceeded 1. For the 500 MLE's, the mean and median both were close to the true value of PTE (0.7). Over 10% of the MLE's for PTE were outside of the unit interval. For the diffuse normal prior analysis,

over 22.4% of the posterior means were outside the unit interval. However, posterior means and medians were close to the true *PTE*.

By construction, all Bayesian point and interval estimates were within the unit interval. In addition, there was less variability among the posterior estimates for *PTE*. Both priors were somewhat pessimistic in this example (means and medians less than 0.7, with modes equal to 0.7) so, not surprisingly, Bayesian point estimates somewhat underestimate the *PTE*.

PTE = 0.9 In this scenario, we generate data using $\beta_\gamma = 0.375$ and $\gamma = 3.375$, yielding *PTE* = 0.9. In the MCMC simulation we used two chains with 1,000 burn-in iterations followed by 20,000 sample iterations. We encountered serious convergence problems that are discussed in the Appendix A. Parameters $(\mu_\gamma, \mu, \gamma)$ were assumed to follow independent $N(0, 10)$ priors instead of the diffuse normal priors used earlier. Table 5 displays averages of mean, median and 95% intervals of *PTE* estimations with four prior structures used with the generated 500 data sets.

MLE's for *PTE* were close to the true value of *PTE* (0.9) and exhibited less variability in the 500 data sets compared to the results with $N = 40$. About 4% of MLE's *PTE* are greater than 1.

With the diffuse normal prior structure, the posterior means for *PTE* were far from the true value *PTE* and a substantial number were outside the unit interval. Equal-tail credible sets based on the diffuse normal prior structure were equally poor and over 20% of the corresponding posterior means for *PTE* fell outside $[0,1]$.

Table 5. Simulation result with four prior structures when true $PTE = 0.9$ and $N = 40$. The average of the 500 MLE's was 0.85 with 2.5% and 97.5% quantiles 0.676 and 1.055, respectively. The median of the MLE's was 0.890. Table entries summarize posterior features for the Bayesian analyses. Parenthetic entries are the 2.5% and 97.5% quantiles for 500 values.

Method	95% equal tail credible interval			
	mean	2.5%	median	97.5%
Diffuse	0.298	-2.416	0.465	1.875
Normal	(-4.982, 3.438)	(-17.357, 0.359)	(-0.373, 1.660)	(0.748, 20.030)
Uniform	0.853 (0.608, 0.938)	0.655 (0.230, 0.816)	0.863 (0.630, 0.950)	0.988 (0.866, 0.998)
Beta Prior 1	0.784 (0.687, 0.877)	0.576 (0.467, 0.669)	0.791 (0.698, 0.884)	0.945 (0.847, 0.977)
Beta Prior 2	0.765 (0.640, 0.891)	0.565 (0.384, 0.768)	0.773 (0.647, 0.898)	0.915 (0.827, 0.977)

Bayesian analysis using the conditional uniform prior structure yields posterior means and medians for PTE close to the true value (0.9). Again, because the beta 1 and beta 2 priors are somewhat pessimistic, and since the sample is relatively small, posterior medians underestimated PTE . However, credible sets based on either the beta 1 or beta 2 prior were somewhat narrower than in the previous case. It must be remembered, however, that the normal components of the joint prior had smaller variances here.

3.2.3 Moderate Sample Size ($N = 200$)

In this section we use the same prior structures as in Section 3.2.2 but we increase the sample size to 200. We generated 100 data sets, each with 200 patients, equally divided into two treatment arms. As before, Bayesian analyses are conducted in WinBUGS. For convergence diagnostic purposes, we computed Gelman-Rubin statistics and examined autocorrelation plots. There is no convergence problem presents in the first two scenarios. Specifically, in the first two scenarios ($PTE = 0.5$ and 0.7), MCMC

simulations were based on two chains, beginning with a burn-in sample of 1,000 and a subsequent sample based on 10,000 iterations. In the third scenario ($PTE = 0.9$) we again encountered convergence problems. As before, we increase the number of iterations from 10,000 to 20,000 with 1,000 burn-in iterations for two chains. We retained every 10th iteration after burn-in and discarded the rest. The normal components of the joint prior structure were given variance 10, as in the case for $PTE = 0.9$ and $N = 40$.

$PTE = 0.5$ In this scenario, we generated 100 data sets each with 40 triples (E_i, T_i, S_i) using $\beta_\gamma = 0.75$ and $\gamma = 0.375$, yielding $PTE = 0.5$. Table 6 shows averages of mean, median and 95% intervals of PTE with the four priors described in Section 3.2.1.

Table 6. Simulation results with four prior structures when true $PTE = 0.5$ and $N = 200$. The average of the 100 MLE's was 0.483 with 2.5% and 97.5% quantiles 0.0872 and 0.926, respectively. The median of the MLE's was 0.481. Table entries summarize posterior features for the Bayesian analyses. Parenthetic entries are the 2.5% and 97.5% quantiles for 100 values.

Method	95% equal tail credible interval			
	mean	2.5%	median	97.5%
Diffuse Normal	0.465 (0.0482, 0.922)	-0.197 (-0.886, 0.276)	0.484 (0.0896, 0.928)	0.999 (0.604, 1.581)
Uniform	0.484 (0.250, 0.716)	0.0486 (0.0117, 0.205)	0.487 (0.223, 0.755)	0.918 (0.638, 0.987)
Beta Prior1	0.550 (0.350, 0.724)	0.161 (0.0719, 0.323)	0.555 (0.339, 0.752)	0.905 (0.702, 0.975)
Beta Prior2	0.610 (0.468, 0.720)	0.306 (0.204, 0.429)	0.614 (0.465, 0.734)	0.879 (0.741, 0.936)

Here the MLE's for PTE were inside of the unit interval but varied widely, albeit not as much as with $N = 40$. About 3% of MLE's for PTE were outside of $[0,1]$. The diffuse normal prior-based analysis performed poorly, with at least 3% of the posterior means being outside the unit interval.

The posterior means and medians from the Bayesian analyses are closest to the true value under the conditional uniform prior structure. What is striking is the relatively small improvement in credible set width over the smaller sample case—compare to the results in Table 3. Indeed, we saw little improvement in credible set width until historically driven informative priors were used, as will be seen in Section 3.3.

$PTE = 0.7$ In this scenario, we generated 100 data sets each with 40 triples (E_i, T_i, S_i) using $\beta_\gamma = 0.375$ and $\gamma = 0.875$, yielding $PTE = 0.7$. Table 7 shows averages of mean, median and 95% intervals for PTE under the four prior structures based on 100 samples when $PTE = 0.7$ and $N = 200$.

Table 7. Simulation results with four prior structures when true $PTE = 0.7$ and $N = 200$. The average of the 100 MLE's was 0.711 with 2.5% and 97.5% quantiles 0.481 and 0.986, respectively. The median of the MLE's was 0.702. Table entries summarize posterior features for the Bayesian analyses. Parenthetic entries are the 2.5% and 97.5% quantiles for 100 values.

Method	95% equal tail credible interval			
	mean	2.5%	median	97.5%
Diffuse Normal	0.697 (0.474, 0.984)	0.167 (0.134, 0.201)	0.364 (0.482, 0.984)	1.004 (0.759, 1.336)
Uniform	0.679 (0.466, 0.838)	0.349 (0.130, 0.552)	0.687 (0.472, 0.861)	0.947 (0.752, 0.993)
Beta Prior1	0.682 (0.502, 0.820)	0.388 (0.215, 0.552)	0.690 (0.506, 0.839)	0.928 (0.760, 0.983)
Beta Prior2	0.685 (0.547, 0.787)	0.438 (0.314, 0.557)	0.692 (0.549, 0.799)	0.894 (0.769, 0.950)

In contrast to small sample results presented in Table 4, the MLE's in this case were much closer to the true value of PTE (0.7) and have less variability across the 100 samples. Furthermore, only one out of 100 data sets resulted in a negative MLE for PTE .

Once again, the diffuse normal prior approach faired poorly. The posterior medians with diffuse normal priors were far from the true value (0.7) although its posterior means were close to 0.7. About 18 out of 100 97.5% percentiles for posterior PTE 's were greater than 1.

The posterior means and medians using the Bayesian analysis with conditional priors are much closer to the true value of PTE (0.7), even for the conditional uniform priors, which were centered on 0.5. Credible set widths were narrower than for the $N = 40$ case but, again, the improvement was not dramatic. Indeed, even increasing the sample size to $N = 500$ has little effect. To illustrate this, we generated 500 observations from the same model and used the same prior structure. We computed $\widehat{PTE} = 0.6855$. Using WinBUGS we ran parallel chains with the same burn-in and sample length as described above. The posterior mean for PTE was 0.6811. The posterior 95% equal-tail credible set for PTE was $[0.4619, 0.8828]$, with a median of 0.683. Note that the credible set width is not much smaller than those in the simulation summarized in Table 4.

$PTE = 0.9$ In this scenario, we generated 100 data sets with $\beta_\gamma = 0.375$ and $\gamma = 3.375$, yielding $PTE = 0.9$. Table 8 shows averages of mean, median and 95% intervals of PTE with four prior structures from 100 samples when $PTE = 0.9$ and $N = 200$.

Compared with Table 5, the posterior means and medians from Bayesian analysis are much closer to the true value of PTE (0.9). The posterior PTE is not dramatically affected by the pessimistic prior distribution. Once again, analysis using diffuse normal distributions is poor. About 35% of the upper bounds for 95% credible sets for posterior PTE are greater than 1. All other methods performed relatively well in this scenario, including maximum likelihood. Note that 95% quantiles for posterior means, medians,

and credible set endpoints are much narrower compared to Table 5. In general, point and interval estimates for endpoints exhibited less variability in this scenario.

Table 8. Simulation results with four prior structures when true $PTE = 0.9$ and $N = 200$. The average of the 100 MLE's was 0.894 with 2.5% and 97.5% quantiles 0.821 and 0.966, respectively. The median of the MLE's was 0.893. Table entries summarize posterior features for the Bayesian analyses. Parenthetic entries are the 2.5% and 97.5% quantiles for 100 values.

Method	95% equal tail credible interval			
	mean	2.5%	median	97.5%
Diffuse Normal	0.891 (0.819, 0.965)	0.815 (0.735, 0.898)	0.892 (0.820, 0.965)	0.962 (0.896, 1.038)
Uniform	0.888 (0.819, 0.953)	0.815 (0.733, 0.892)	0.888 (0.820, 0.956)	0.960 (0.896, 0.997)
Beta Prior1	0.886 (0.815, 0.946)	0.812 (0.730, 0.887)	0.887 (0.817, 0.949)	0.954 (0.893, 0.993)
Beta Prior2	0.875 (0.811, 0.933)	0.803 (0.725, 0.876)	0.877 (0.812, 0.935)	0.941 (0.886, 0.980)

3.2.4 Conclusions

The following conclusions are suggested by the small simulation studies summarized in the preceding sections.

1. Point estimates for all methods were reasonably close to the true value of PTE , except for those derived using a Bayesian analysis with diffuse normal priors in the case of $PTE = 0.9$.
2. All prior structures produced somewhat wide interval estimates, even when $N = 200$.
3. Use of the MLE or a Bayesian analysis with diffuse normal priors is not advised.

Estimates of PTE using these methods are routinely outside of the unit interval.

4. Increasing the sample size from 40 to 200 had, at best, only a moderate payoff in reducing interval widths. Single samples of size 500 and 1000 produced only modest improvement.

3.3 Informative Priors from Historical Data

We have already noted that large randomized trials or meta-analysis of multiple trials will often be required for precise estimation of *PTE* (Molenberghs et al., 2005). A possible alternative to either of these approaches is to use an appropriate historical study to fit conditional beta prior distributions. This approach begins with relatively non-informative priors on all parameters. A joint posterior is derived using the historical data. This posterior becomes the prior for use with the current data. We consider such priors in this section. Related prior based on historical information, the so-called “power prior”, is examined in Section 5.3.

3.3.1 Posterior Using Priors Based on Historical Data

We simulated two historical data sets, for sample sizes of $N_h = 20$ and $N_h = 100$. The surrogate endpoints, S , in both historical data sets were simulated from a normal distribution with variance 2 and means 0.5 and -0.5 for the treatment and control groups, respectively. The true endpoint, E , was simulated from a Poisson distribution with mean

$$\lambda = \exp(\mu_\gamma + \beta_\gamma T + \gamma S).$$

As before, we have assumed models (3.1) and (3.2) hold simultaneously, resulting in the likelihood functions (2.8) and (2.9). Thus, the historical data was generated from the same model ($PTE = 0.7$) used to produce data sets of size 40 and 200 for the simulations in the last section.

To obtain a posterior distribution using the historical data, we used relatively non-informative priors. Thus, coefficient parameters μ , μ_γ , and γ in models (3.1) and (3.2) were assumed to have independent diffuse normal priors, as in Section 3.1. We assumed that β and β_γ have conditional uniform prior distributions; that is, $\beta \sim \text{beta}_{[0, d]}(1, 1)$ and $\beta_\gamma \sim \text{beta}_{[0, \beta]}(1, 1)$. We could have chosen d using a sensitivity analysis as was done in Chapter 2, but we opted instead to use an arbitrarily “large” value, setting $d = 100$.

3.3.2 Prior Structures Based on Historical Data (Sample Size $N_h = 20$ and $N_h = 100$)

Denote by ξ_θ and σ_θ^2 the posterior mean and variance, respectively, of a parameter θ , using the posterior based on historical data. We fit a prior for use with the current data by matching moments. Thus, for example, suppose ξ_{μ_S} and $\sigma_{\mu_S}^2$ are the posterior mean and variance of μ_S using the historical data. Then we use the prior $\mu_S \sim N(\xi_{\mu_S}, \sigma_{\mu_S}^2)$ with the current data. Priors for the parameters γ , μ , and μ_γ are constructed in a similar fashion.

The variance component τ is assumed to follow an inverse-gamma distribution. We construct the prior $\tau^{-1} \sim \text{gamma}(u, v)$ where (u, v) is the solution to the pair of equations $\xi_\tau = uv$ and $\sigma_\tau^2 = uv^2$.

We use the prior $\beta \sim \text{beta}_{[0, 100]}[a, b]$ with the current data, where

$$a = \frac{-\sigma_\beta^2 \mu_\beta}{\sigma_\beta^2 - \mu_\beta + \mu_\beta^2} \quad (3.6)$$

and

$$b = \frac{\sigma_{\beta}^2(\mu_{\beta} - 1)}{\sigma_{\beta}^2 - \mu_{\beta} + \mu_{\beta}^2} \quad (3.7)$$

Similarly, we take $PTE \sim \text{beta}_{[0, 1]}[c, d]$, where

$$c = \frac{-\sigma_{PTE}^2 \mu_{PTE}}{\sigma_{PTE}^2 - \mu_{PTE} + \mu_{PTE}^2} \quad (3.8)$$

and

$$d = \frac{-\sigma_{PTE}^2(\mu_{PTE} - 1)}{\sigma_{PTE}^2 - \mu_{PTE} + \mu_{PTE}^2} \quad (3.9)$$

The corresponding β_{γ} prior structure is $\beta_{\gamma} \sim \text{beta}_{[0, \beta]}[d, c]$.

Table 9 shows the priors constructed for parameters $(\gamma, \mu, \mu_{\gamma}, \mu_s, \tau)$ in model (3.1) and (3.2), based on the historical data with sample size $N_h = 20$ and $N_h = 100$ as described above.

Table 9. Prior structures and 95% intervals for parameters.

Parameters		γ normal	μ normal	μ_{γ} normal	μ_s normal	τ gamma
$N_h = 20$	mean	0.83	1.431	0.176	0.285	9.36
	variance	0.075	0.127	0.226	0.424	29.85
	95% interval	(0.679, 0.973)	(1.165, 1.66)	(0.733, 1.619)	(-0.546, 1.115)	(130.1, 484.8)
$N_h = 100$	mean	0.847	1.53	0.15	-0.36	9.36
	variance	0.025	0.053	0.1	0.23	29.85
	95% interval	(0.798, 0.896)	(1.426, 1.634)	(-0.044, 0.344)	(-0.802, 0.082)	(130.1, 484.8)

Tables 10 and 11 summarize the prior structures for β and β_{γ} constructed using the historical data using beta and normal priors. The former were constructed using (3.6)-(3.9). The latter were constructed by matching moments, much as with the other

model parameters with normal priors in Table 9. We considered both informative normal priors and informative beta priors for β and β_γ . Normal and beta priors were selected to have similar 95% equal-tail intervals for β . Since we constrained β_γ to have support $[0, \beta]$, the 95% equal-tail intervals for β_γ under the normal prior and beta prior structure might be slightly different.

Table 10. Normal prior structures for β , β_γ and their 95% intervals.

		Normal prior	
		β	β_γ
$N_h = 20$	mean	1.57	0.67
	variance	0.25	0.27
	95% intervals	(1.08, 2.07)	(0.138, 1.196)
$N_h = 100$	mean	1.54	0.442
	variance	0.106	0.113
	95% intervals	(1.327, 1.745)	(0.22, 0.663)

Table 11. Conditional beta prior structures for β , β_γ and their 95% Intervals.

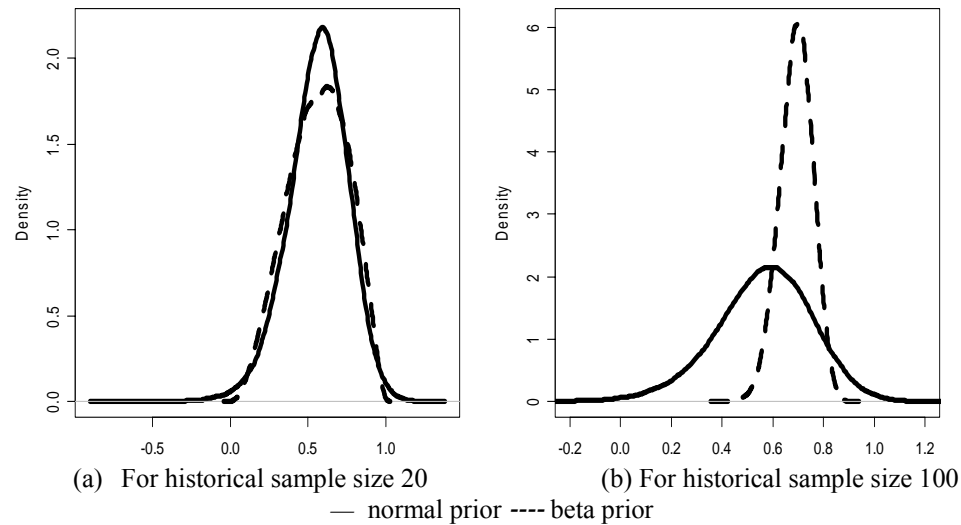
		Conditional Beta prior	
		β	$\beta_\gamma \beta$
$N_h = 20$	support	[0,5]	[0, β]
	shape parameters	(26.11, 56.88)	(2.43, 3.14)
	95% intervals	(1.10, 2.09)	(0.15, 1.38)
$N_h = 100$	support	[0,5]	[0, β]
	shape parameters	(121.16, 281.38)	(15.11, 33.91)
	95% intervals	(1.29, 1.73)	(0.275, 0.686)

Table 12 summarizes the induced prior for *PTE* and its 95% equal-tail intervals. The informative normal prior structure for β and β_γ results in a Cauchy distribution for *PTE*. The informative beta prior for β and β_γ results in a standard beta distribution.

Table 12. Induced prior structures for *PTE* and their 95% Intervals.

		Induced prior for <i>PTE</i>		
	Prior structures for β and β_γ	support	shape parameters	95% intervals
$N_h = 20$	Informative normal	$(-\infty, \infty)$	Cauchy	(0.145, 0.914)
	Informative beta	$[0, 1]$	Beta(3.14, 2.43)	(0.183, 0.903)
$N_h = 100$	Informative normal	$(-\infty, \infty)$	Cauchy	(0.143, 0.912)
	Informative beta	$[0, 1]$	Beta(33.91, 15.11)	(0.558, 0.811)

Figure 9 shows the induced prior for *PTE* using the conditional beta prior method and informative normal prior method, where inferences are from the historical sample with sizes 20 and 100. Figure 9(a) displays the induced priors for *PTE* based on the historical data with sample size 20. Similarly, Figure 9(b) indicates the induced prior for *PTE* based on the historical sample with size 100.

Figure 9. Induced informative priors for *PTE* with normal/beta structure.

Clearly, even with an informative normal structure, the induced prior for *PTE* may have supports outside the unit interval. By construction, the informative conditional

beta prior does not have this problem and therefore takes greater advantage of increased historical sample sizes. This is illustrated in Figure 9(b).

We combined the priors described above with data generated from the same model for “current” sample sizes of $N_c = 40$ and 200. The “current” data with sample size 40 resulted in $\widehat{PTE} = 0.8$ while sample size 200 yielded $\widehat{PTE} = 0.892$.

Table 13 contains the posterior summaries for PTE using informative prior structures. “Normal” and “beta” refer to the priors structures in Tables 10-12.

For the informative normal prior Bayesian analysis, the 97.5 percentile of posterior PTE exceeds 1. The width of the 95% credible sets for PTE using an informative normal prior structure is substantially wider compared with the results using a conditional beta prior structure. As the “current” sample size increases from 40 to 100, the posterior distribution for PTE is dominated by the likelihood. The posterior mean and means for PTE are close to the MLE.

Table 13. Posterior statistics for PTE with informative prior structures.

			mean	2.5%	median	97.5%	Std
$N_h = 20$	$N_c = 40$	normal	0.753	0.47	0.666	1.008	0.137
		beta	0.721	0.480	0.729	0.917	0.113
	$N_c = 200$	normal	0.88	0.766	0.842	0.995	0.058
		beta	0.858	0.753	0.860	0.950	0.05
$N_h = 100$	$N_c = 40$	normal	0.755	0.498	0.76	0.992	0.125
		beta	0.710	0.592	0.712	0.817	0.058
	$N_c = 200$	normal	0.878	0.764	0.878	0.989	0.057
		beta	0.787	0.705	0.788	0.861	0.039

Informative conditional beta priors were compared with conditional uniform priors in Table 14. The results for $N_h = 20$ using the informative conditional beta prior were an improvement over the conditional uniform results for $N_c = 40$ in Table 3.3 with respect to interval width, with $[0.377, 0.945]$ for the former and $[0.0611, 0.974]$ for the latter (on average). Improvement was less dramatic when compared to the conditional uniform prior results for $N = 200$ in Table 3.6: the ($N_c = 20$) informative conditional beta prior yielded $[0.434, 0.957]$ while the conditional uniform prior yielded $[0.349, 0.947]$ (on average) at this sample size.

Table 14. Posterior credible sets for *PTE* using historical data for conditional beta priors compared to the conditional uniform priors. Bounds for the latter are averages as reported in Tables 4 and 7.

N_c	Prior	N_h	lower 95%	upper 95%	width
40	beta	20	0.377	0.945	0.568
		100	0.643	0.854	0.211
		500	0.727	0.907	0.18
		1000	0.747	0.881	0.134
		--	0.0611	0.974	0.913
	uniform	--			
200	beta	20	0.434	0.957	0.523
		100	0.626	0.846	0.220
		500	0.790	0.920	0.130
		1000	0.791	0.896	0.105
		--	0.349	0.947	0.598
	uniform	--			

The results for $N_h = 100$ exhibit improvement to a greater degree over those for the uniform conditional prior summarized in Table 14. As historical sample size increases, the widths for 95% credible sets decreased, as expected. Since we computed prior parameters by matching posterior moments, we did not make use of other features of the “historical” posterior, such as the shape of the distribution. In Chapter 5, we will

discuss a prior method based on historical that can make use of such additional information.

3.4 Sensitivity Analysis

This section, we will discuss the effect of shape parameters used for the β and β_γ conditional beta prior distributions described in Section 3.2.1, that is,

$$\beta \sim \text{beta}_{[0, d]}(a, b),$$

and

$$\beta_\gamma | \beta \sim \text{beta}_{[0, \beta]}(f, g).$$

Specifically, we examine the effect of a , b , f , and g on posterior estimations of PTE . In this sensitivity analysis, we arbitrarily set $d = 100$. We chose four sets of values for the β priors and four sets for the β_γ priors:

$$(a, b) = (15, 46), (36, 46), (15, 66), (36, 66) \tag{3.10}$$

$$(f, g) = (1.5, 2), (3.5, 2), (1.5, 4), (3.5, 4)$$

These shape parameters were selected to complement the informative ($N_h = 20$) conditional beta prior used in the last section. The corresponding induced priors on PTE are summarized in Table 15 and graphed in Figure 10.

In this study, we used the “current” data with size 40 and size 200 that were simulated in Section 3.3, in which $PTE = 0.7$. We had $\widehat{PTE} = 0.799$ for $N_c = 40$ and 0.826 for $N_c = 200$. For each of the 16 combinations of shape parameters, we conducted MCMC simulations using WinBUGS. In each case we used two chains with dispersed initial values beginning with 1,000 burn-in iterations and continued with 10,000 sample

iterations. No convergence problems were encountered. The 16 sets of prior structures are shown in the Table 16 along with corresponding posterior median *PTE* and 95% credible sets.

Table 15. Summary of induced *PTE* priors using the conditional beta priors in (3.10).

Prior	$\text{beta}(g, f)$	mean	median	mode	95% interval
1	(2, 1.5)	0.571	0.586	0.667	[0.118, 0.953]
2	(2, 3.5)	0.364	0.346	0.286	[0.059, 0.759]
3	(4, 1.5)	0.727	0.756	0.545	[0.33, 0.975]
4	(4, 3.5)	0.533	0.756	0.857	[0.202, 0.848]

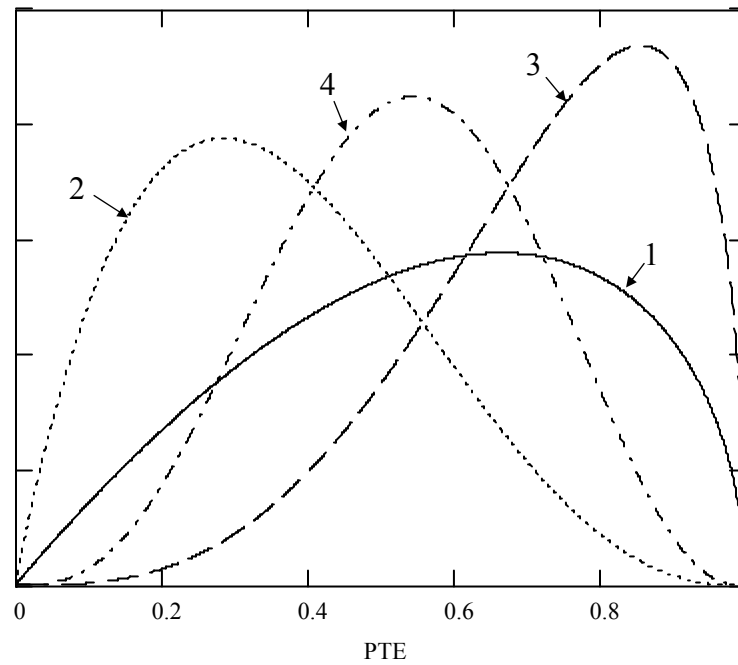


Figure 10. Induced *PTE* priors using conditional beta priors in (3.10).

The change of shape parameters for the conditional β , β_γ priors did not dramatically affect the posterior distribution for *PTE*. Figure 11 displays the posterior

for PTE for four sets of β_γ prior structures when the prior for β is fixed with shape parameters 15 and 46. These were typical results. As can be seen, some of the posterior modes and medians exceeded the true value of 0.7, however, all interval estimates contained 0.7. Given the relatively wide range of priors considered, it appears the posteriors are relatively robust.

Table 16. Posterior median and 95% credible set for PTE for different combinations of prior for β and β_γ with sample size 40.

	$\text{beta}_{[0, 100]}(15, 46)$	$\text{beta}_{[0, 100]}(36, 46)$	$\text{beta}_{[0, 100]}(15, 66)$	$\text{beta}_{[0, 100]}(36, 66)$
$\text{beta}_{[0, \beta]}(1.5, 2)$	0.794 (0.554, 0.961) ¹	0.832 (0.645, 0.968)	0.796 (0.556, 0.962)	0.835 (0.644, 0.969)
$\text{beta}_{[0, \beta]}(3.5, 2)$	0.698 (0.456, 0.877) ²	0.758 (0.567, 0.9)	0.701 (0.458, 0.88)	0.759 (0.568, 0.9)
$\text{beta}_{[0, \beta]}(1.5, 4)$	0.825 (0.605, 0.97) ³	0.792 (0.672, 0.974)	0.823 (0.607, 0.969)	0.851 (0.67, 0.974)
$\text{beta}_{[0, \beta]}(3.5, 4)$	0.734 (0.517, 0.897) ⁴	0.849 (0.672, 0.974)	0.733 (0.516, 0.894)	0.776 (0.595, 0.912)

Note: the number on the upper corner of parenthesis in the first column denoted the number of prior structures in the figure 10.

We repeated the robustness study with a sample of size 200 for which we computed $\widehat{PTE} = 0.826$. Since Table 16 indicates that slight changes in the prior structure for β had little effect on the posterior, we only considered two options for β priors in this study: $\text{beta}_{[0, 100]}(15, 46)$ and $\text{beta}_{[0, 100]}(66, 36)$. Table 17 lists the posterior median PTE and its 95% credible sets with sample size 200. Not surprisingly, with this increased sample size, posterior PTE inferences are more robust to the change of shape parameters in prior for β_γ . The 95% credible sets are narrower compared with those in Table 17.

Table 17. Posterior median *PTE* and 95% credible set for *PTE* for different combination of priors for β and β_γ with sample size 200.

	$\text{beta}_{[0, 100]}(15, 46)$	$\text{beta}_{[0, 100]}(66, 36)$
$\text{beta}_{[0, \beta]}(1.5, 2)$	0.818 (0.677, 0.945) ¹	0.852 (0.735, 0.955)
$\text{beta}_{[0, \beta]}(3.5, 2)$	0.771 (0.633, 0.893) ²	0.811 (0.699, 0.910)
$\text{beta}_{[0, \beta]}(1.5, 4)$	0.830 (0.695, 0.952) ³	0.857 (0.744, 0.959)
$\text{beta}_{[0, \beta]}(3.5, 4)$	0.784 (0.678, 0.900) ⁴	0.820 (0.706, 0.917)

Note: the number on the upper corner of parenthesis in the first column denoted the number of prior structures in the figure 10

Figure 12 exhibits the posterior for *PTE* when β_γ has four prior structures and one fixed β prior structure ($\text{beta}_{[0, 100]}(15, 46)$). With this moderate sample size, the posterior *PTE* are dominated by the likelihood. Differences between posterior *PTE*'s with these four prior structures are not as large as those in Figure 11. In short, with moderate sample size ($N_c = 200$), a slight change in prior parameters will not profoundly influence the posterior of the *PTE*.

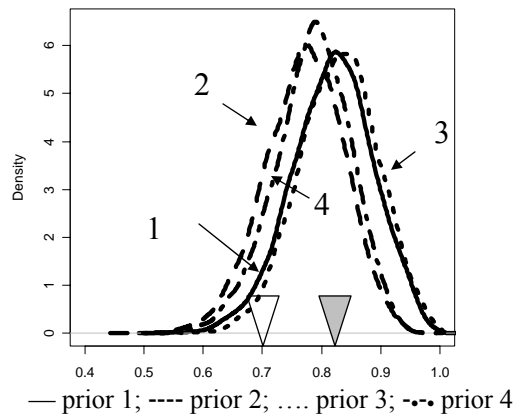


Figure 12. Posterior density of *PTE* with four different β_γ priors and one fixed β prior structure. Numbers refer to the priors labeled in Table 17. The grey triangle marks the MLE, $\widehat{PTE} = 0.826$. The open triangle marks the true value, $PTE = 0.7$.

CHAPTER FOUR

The Effect of Interaction on the Estimation of PTE

This chapter focuses on modeling the uncertainty about interaction effects. In Section 4.1, we discuss the impact of interaction on *PTE* estimation. In Section 4.2 we use one example to illustrate modeling the uncertainty about the interaction effect. More examples are used to illustrate the influence of interaction in Section 4.3. In Section 4.4 we attempt to characterize the effect interaction has on *PTE* estimation. A joint model of interaction and *PTE* are developed in Section 4.5. There we recommend modeling the interaction effect rather than assuming its absence after a hypothesis test. We utilize Bayesian model comparison methods in Section 4.6 to compare models with and without interaction.

4.1 The Effect of Assuming Interaction Is Absent When It Is Not

Suppose we choose to leave the interaction term out of the model when, in fact, interaction is present in the data. How will this situation affect estimation of *PTE*?

Recall from Section 2.1 that $PTE = 1 - \beta_\gamma/\beta$, where the coefficients are from the two linear models

$$\mathcal{E}(E|T, S) = \mu_\gamma + \beta_\gamma T + \gamma S \quad (4.1)$$

and

$$\mathcal{E}(E|T) = \mu + \beta T. \quad (4.2)$$

If, in fact, there is interaction, the model should be

$$\mathcal{E}(E | T, S) = \mu'_\gamma + \beta'_\gamma T + \gamma' S + \eta TS \quad (4.3)$$

Note that if (4.3) is the true model ($\eta \neq 0$) then we cannot speak of the “true” value of *PTE* as it is undefined. Suppose we mistakenly attribute the interaction to the treatment. That is, (4.3) becomes

$$\mathcal{E}(E | T, S) = \mu'_\gamma + (\beta'_\gamma + \eta S)T + \gamma' S \quad (4.4)$$

and, implicitly, we assume that for all S ,

$$\beta'_\gamma + \eta S = \beta_\gamma.$$

Clearly, if $\beta'_\gamma \approx \beta_\gamma$ and η/β is small, we can write $PTE(S) \approx PTE$ for all S . In general, however, *PTE* becomes a function of S :

$$PTE(S) = 1 - \frac{\beta'_\gamma + \eta S}{\beta} = 1 - \frac{\beta'_\gamma}{\beta} - \frac{\eta}{\beta} S. \quad (4.5)$$

Therefore, *PTE* will be under-estimated when $\eta S > 0$ and over-estimated when $\eta S < 0$.

Define

$$h(S) = \frac{\eta}{\beta} S. \quad (4.6)$$

If $h(S)$ is small enough, it is reasonable to proceed as if interaction effect is absent. However, in order for $h(S)$ to be small *for all* S , the variability of S must be minimal. This is itself problematic, since it brings into question whether a biomarker with such small variability could conceivably be a reasonable surrogate. We return to this issue in Section 4.3.

If interaction is attributed to the surrogate, the situation is somewhat similar. However, the effect on estimation of *PTE* is not as direct, as the latter involves β and β_γ , not γ .

4.2 The Influence of Interactions on *PTE* Estimation

Given the problematic nature of interaction, statistical analyses of surrogate endpoints typically begin by testing the hypothesis $H_0: \eta = 0$, in model (4.3). If this hypothesis is not rejected, the term is dropped from the model and the surrogate analysis proceeds under the assumption of no interaction. For instance, Cummings, Black and Thompson (1998) used a proportional hazards ratio model that included treatment on the risk of clinical fractures, BMD, and treatment-by-BMD interaction. The authors tested the significance of the interaction term and failed to reject the null hypotheses. Thereafter they concluded that there is no interaction effect between surrogate endpoint and treatment.

Of course, observing a p -value larger than some specified α for $H_0: \eta = 0$ does not imply that $\eta = 0$. Is it possible that statistically insignificant interaction can have a practical impact on *PTE* estimation? To address this question, we can retain interaction in the model and investigate it jointly with *PTE*. In this section, we present an example that motivates this approach.

For our example, we simulated E_1, \dots, E_n iid values of the true endpoint from a Poisson distribution with mean

$$\lambda = \exp\{\beta'_\gamma T + \gamma' S + \eta TS\}.$$

Here T denotes the treatment:

$$T = \begin{cases} 0.5 & \text{treatment} \\ -0.5 & \text{control}, \end{cases}$$

and S denotes the surrogate, where

$$S|T = -0.5 \sim N(-0.5, 0.5)$$

and

$$S|T = 0.05 \sim N(0.5, 0.5).$$

We generated $n = 40$ triples (E_i, T_i, S_i) using $\beta'_\gamma = 0.36$, $\gamma' = 0.7$, and $\eta = 0.2$ using SAS. The program is included in Appendix B.1. We used SAS PROC GENMOD to fit the Poisson regression of $E|T, S, T \times S$. Poisson regressions for models $E|T, S$ and $E|T$ were fit simultaneously to compute maximum likelihood estimators (MLE's) for β and β_γ . The SAS program is included in Appendix C.1. The MLE for PTE is the ratio of MLE's for β_γ and β in the models (4.1) and (4.2):

$$\widehat{PTE} = 1 - \frac{\hat{\beta}_\gamma}{\hat{\beta}} \quad (4.7)$$

The frequentist test of $H_0: \eta = 0$ yielded a p -value of 0.31 and was therefore not rejected.

In our Bayesian analysis of this simulated data set, we used the same data model described above. Let $\mathbf{E} = (E_1, \dots, E_n)$, $\mathbf{S} = (S_1, \dots, S_n)$, and $\mathbf{T} = (T_1, \dots, T_n)$. Under model (4.1) the likelihood is

$$l_1(\mu_\gamma, \beta_\gamma, \gamma, \eta | \mathbf{E}, \mathbf{S}, \mathbf{T}) = \sum_{i=1}^n [-\mu_{1i} + E_i \log(\mu_{1i}) - \log(E_i)]$$

where

$$\mu_{1i} = \mu_\gamma + \beta_\gamma T_i + \gamma S_i.$$

Under model (4.2) it is

$$l_2(\mu, \beta \mid \mathbf{E}, \mathbf{T}) = \sum_{i=1}^n [-\mu_{2i} + E_i \log(\mu_{2i}) - \log(E_i)]$$

where $\mu_{2i} = \mu + \beta T_i$.

Under model (4.3) the likelihood function is

$$l_3(\mu'_\gamma, \beta'_\gamma, \gamma', \eta \mid \mathbf{E}, \mathbf{S}, \mathbf{T}) = \sum_{i=1}^n [-\mu_{3i} + E_i \log(\mu_{3i}) - \log(E_i)]$$

where $\mu_{3i} = \mu'_\gamma + \beta'_\gamma T_i + \gamma' S_i + \eta T_i S_i$.

We used the following prior structure on the parameters in models (4.1), (4.2), and (4.3). We placed the diffuse normal distribution, $N(0, 100,000)$ on each of μ , μ'_γ , β'_γ , and η . Following the development in Section 2.4, we used the following priors:

$$\beta \sim \text{beta}_{[0,d]}(2,2),$$

$$\beta_\gamma \mid \beta \sim \text{beta}_{[0,\beta]}(2,2),$$

and

$$\beta'_\gamma \mid \beta \sim \text{beta}_{[0,\beta]}(2,2).$$

We chose $d = 5$ using a sensitivity analysis as described in Section 2.5. The joint posterior distributions for the parameters in models (4.1), (4.2), and (4.3) are:

$$p_1(\mu_\gamma, \beta_\gamma, \gamma, \beta \mid \mathbf{E}, \mathbf{S}, \mathbf{T}) \propto l_1(\mu_\gamma, \beta_\gamma, \gamma, \beta \mid \mathbf{E}, \mathbf{S}, \mathbf{T}) p(\beta_\gamma \mid \beta) p(\beta) p(\mu_\gamma) p(\gamma),$$

$$p_2(\mu, \beta \mid \mathbf{E}, \mathbf{T}) \propto l_2(\mu, \beta \mid \mathbf{E}, \mathbf{T}) p(\mu) p(\beta)$$

and

$$p_3(\mu'_\gamma, \beta'_\gamma, \gamma', \eta, \beta \mid \mathbf{E}, \mathbf{S}, \mathbf{T}) \propto l_3(\mu'_\gamma, \beta'_\gamma, \gamma', \eta \mid \mathbf{E}, \mathbf{S}, \mathbf{T}) p(\beta'_\gamma \mid \beta) p(\beta) p(\mu'_\gamma) p(\gamma') p(\eta)$$

The marginal posterior for η is given by

$$p(\eta | \mathbf{E}, \mathbf{S}, \mathbf{T}) = \int \int \int \int p_3(\mu'_\gamma, \beta'_\gamma, \gamma', \eta, \beta | \mathbf{E}, \mathbf{S}, \mathbf{T}) d\mu'_\gamma d\beta d\beta'_\gamma d\gamma'$$

where the integration is over all possible values of μ'_γ , β , β'_γ , and γ' . The posterior for *PTE* is the distribution of $1 - \beta_\gamma/\beta$ derived from the joint posterior of (β_γ, β) .

There are no closed-form expressions for the posterior distributions described above, so we used MCMC methods as implemented in WinBUGS. The WinBUGS model and R-code to call WinBUGS are very similar to Appendix D.2 and E.1. Beginning two chains with over-dispersed starting values, we discarded the first 1,000 iterations and based our posterior computations on a further 10,000 iterations. For convergence diagnostic purposes, we computed Gelman-Rubin statistics and examined autocorrelation plots. We found no problems with convergence of the chains.

Figure 13 displays the posterior distribution for the interaction effect, η . The posterior mode for η is 0.45 and $P(\eta > 0 | \mathbf{E}, \mathbf{T}, \mathbf{S}) = 0.7885$.

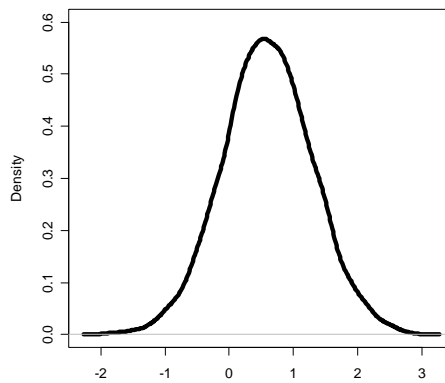


Figure 13. Posterior Density for η .

Clearly one should not simply assume that there is no interaction effect, a p -value in excess of 0.05 notwithstanding. This is made further evident by looking at $h(S) = \eta S / \beta$ from (4.6). One can see from Figure 14 that $h(S)$ is in no way sufficiently small to ignore.

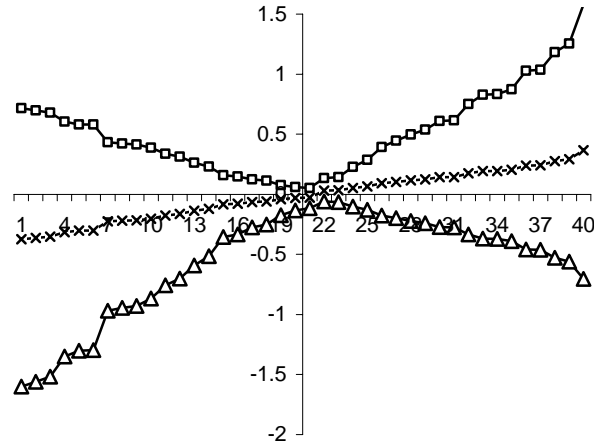


Figure 14. Median and 95% credible set values for $h(S)$. The abscissa is S . Median, upper, and lower bounds are indicated with Δ , \diamond , and \square respectively.

As this example illustrates, ignoring interaction, even after failing to reject the appropriate frequentist hypothesis test, can result in misleading inferences about PTE . In the next section we further investigate the role of interaction in such inference.

4.3 The Affect of Interaction on PTE Estimation

In this section, we will use three scenarios to demonstrate further the effect of interaction on PTE estimations. In each scenario, single data sets of size 40 and 200 were simulated. In each data set, half of the simulated observations were assigned to the treatment group and half to the control group. We used the Poisson-normal model

described in the last section for the true effect and surrogate values. We included an interaction term, as in model (4.3). In each scenario we used $\beta'_\gamma = 0.36$ and $\gamma' = 0.7$. The scenarios differ in their value of η . We used $\eta = 0.2, 1$, and 1.5 .

In each scenario, MCMC methods were used to compute posterior quantities of interest. We used 10,000 iterations from two chains after discarding 1,000 burn-in iterations. For convergence diagnostic purposes, we computed Gelman-Rubin statistics and examined autocorrelation plots. There were no problems with convergence of the chains.

As in the last section, we placed the diffuse normal distribution, $N(0, 100,000)$ on each of μ , μ_γ , μ'_γ , and η . We used the conditional priors

$$\beta \sim \text{beta}_{[0,h]}(a, b),$$

$$\beta_\gamma | \beta \sim \text{beta}_{[0,\beta]}(c, d),$$

and

$$\beta'_\gamma | \beta \sim \text{beta}_{[0,\beta]}(c, d).$$

for various choices of a , b , c , and d . We again chose $h = 5$ based on a sensitivity analysis like that described in Section 2.5. Note that we used the same conditional prior for β_γ and β'_γ . Since we have no prior information on PTE , four prior structures are used for β and β_γ --uniform, symmetric unimodal, left-skewed, and right-skewed. These priors and the consequent induced priors on PTE are detailed below:

$$(1) \text{ Prior 1: } \beta \sim \text{beta}_{[0,5]}(2, 2) \text{ and } \beta_\gamma \sim \text{beta}_{[0,\beta]}(2, 2) \Rightarrow PTE \sim \text{beta}_{[0,1]}(2, 2)$$

$$(2) \text{ Prior 2: } \beta \sim \text{beta}_{[0,5]}(2, 2) \text{ and } \beta_\gamma \sim \text{beta}_{[0,\beta]}(2, 5) \Rightarrow PTE \sim \text{beta}_{[0,1]}(5, 2)$$

$$(3) \text{ Prior 3: } \beta \sim \text{beta}_{[0,5]}(2, 2) \text{ and } \beta_\gamma \sim \text{beta}_{[0,\beta]}(5, 2) \Rightarrow PTE \sim \text{beta}_{[0,1]}(2, 5)$$

$$(4) \text{ Prior 4: } \beta \sim \text{beta}_{[0,5]}(1, 1) \text{ and } \beta_\gamma \sim \text{beta}_{[0,\beta]}(1, 1) \Rightarrow PTE \sim \text{beta}_{[0,1]}(1, 1)$$

The induced priors on PTE were derived using the results of Section 2.4. These priors for PTE are graphed in Figure 15.

In these examples, posterior PTE 's were generated in WinBUGS from paired posterior values of β and β_γ based on equation (4.1) and (4.2), as described in the last section, using posteriors $p_1(\mu_\gamma, \beta_\gamma, \gamma, \beta | \mathbf{E}, \mathbf{S}, \mathbf{T})$ and $p_2(\mu, \beta | \mathbf{E}, \mathbf{T})$. The posterior for η is computed using WinBUGS samples from the posterior $p_3(\mu'_\gamma, \beta'_\gamma, \gamma', \eta, \beta | \mathbf{E}, \mathbf{S}, \mathbf{T})$.

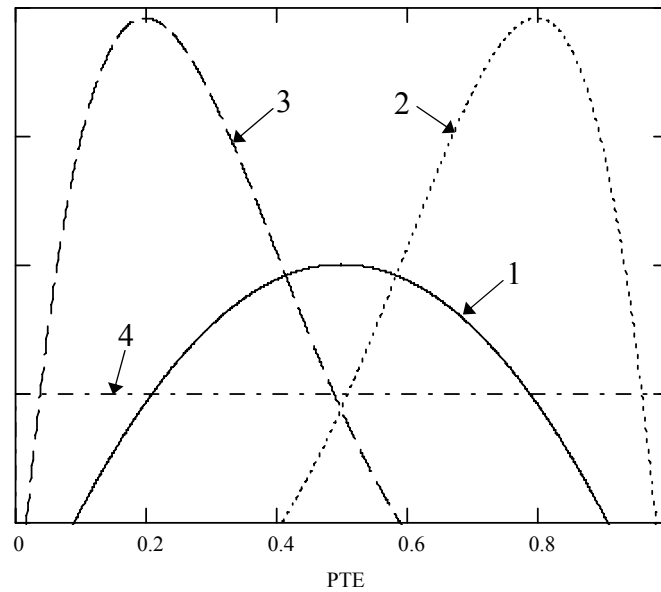


Figure 15. Induced prior for PTE with four prior structures. Numbers refer to the prior descriptions in the text.

4.3.1 Scenario 1: Small Interaction ($\eta = 0.2$)

Table 18 illustrates frequentist estimators and posterior median estimators for PTE and the interaction effect η . The frequentist hypothesis testing p -value for each data set is also shown. The hypothesis of no interaction would not be rejected in either case.

Interval estimates for the small sample case are rather wide, as was the case with many of the examples in Chapters 2 and 3. Furthermore, the priors have considerable

influence over the posterior for PTE . Finally, the MLE and posterior medians are hardly in agreement. These facts reflect the fact that PTE is based on parameters from two weakly linked models, (4.1) and (4.2). We shall see in Chapter 5 that prior information on β and β_γ is critical if reliable estimation of PTE is to be possible. Interval estimates are narrower in the larger sample case, as expected, as is agreement with the MLE. Also, as expected, the posterior for PTE in the larger sample case is not as influenced by the prior structure.

Table 18. Estimators for PTE and η when true $\eta = 0.2$. (Recall there is no “true” value of PTE since $\eta \neq 0$.) Lower (upper) refers to the 2.5th (97.5th) percentile of the posterior. Bayesian point estimates are posterior medians. For the case of $N = 40$, the MLE’s for PTE and η are 0.195 and 0.809, respectively. The values are 0.519 and -0.109 in the case of $N = 200$. The p -values are for testing the hypothesis of no interaction.

$\eta = 0.2$								
	$N = 40$				$N = 200$			
	Prior 1	Prior 2	Prior 3	Prior 4	Prior 1	Prior 2	Prior 3	Prior 4
PTE	0.525	0.280	0.752	0.356	0.511	0.631	0.367	0.628
Lower	0.103	0.047	0.376	0.021	0.190	0.363	0.101	0.368
Upper	0.918	0.66	0.960	0.875	0.799	0.869	0.647	0.867
η	0.423	0.427	0.426	0.581	-0.114	-0.129	-0.076	-0.133
Lower	-0.886	-0.869	-0.849	-0.824	-0.722	-0.707	-0.697	-0.718
Upper	1.772	1.758	1.756	1.988	0.479	0.462	0.523	0.447
p -value	0.31				0.725			

Bayesian inference for η is largely independent of priors on the other parameters. The Bayesian point and interval estimates for the small sample case underestimate η . In the larger sample case, the true value of η is included in the interval estimates for all four prior structures. The MLE for η is poor in both the small and the larger sample case.

Figure 16 compares the posterior kernel density plots for PTE and η using prior1 with two different sample sizes (40 and 200).

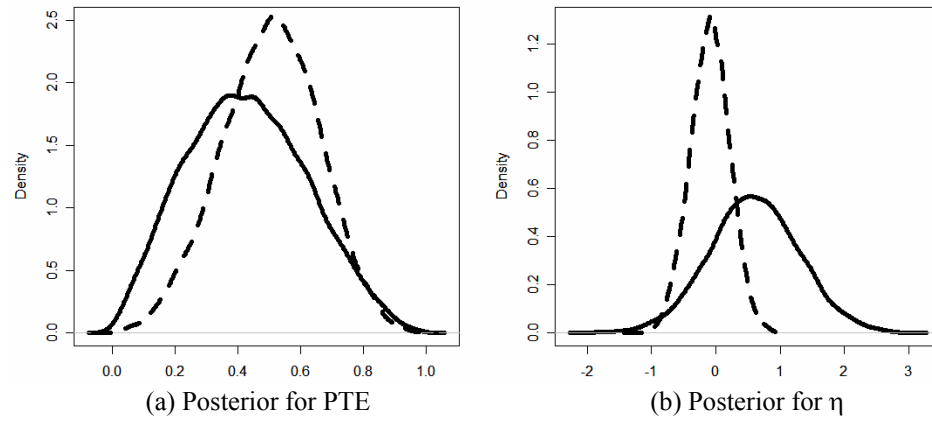


Figure 16. Posterior densities for PTE and η with sample size 40 and 200 respectively using Prior 1. The true value is $\eta = 0.2$. Here the solid line represents posteriors with $N = 40$ while the dashed line is for $N = 200$.

Figure 17 shows the posterior for η with three prior structures when the sample size is 40. There is even greater agreement for the case of $N = 200$. Although the priors differ for the other parameters, the same diffuse normal is used for all four joint prior structures, so this agreement is expected.

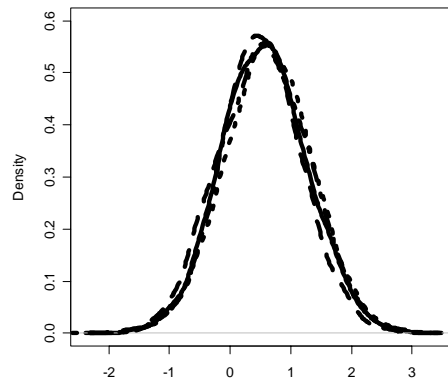


Figure 17. Posterior densities for the interaction using three prior structures for $N = 40$. The true value is $\eta = 0.2$. Here the solid line corresponds to prior 1, the dashed line to prior 2, dotted line to prior 3, and the dot-dash line to prior 4.

4.3.2 Scenario 2 ($\eta = 1$)

Table 19 contains frequentist estimators and posterior median estimators for *PTE* and η . The frequentist hypothesis testing *p*-values are for testing the hypothesis of no interaction. The hypothesis of no interaction would be rejected in the larger sample case and not rejected in the small sample case. For the case of $N = 40$, the MLE's for *PTE* and η are 0.435 and 1.231, respectively. The values are 0.744 and 0.851 when $N = 200$.

Table 19. Estimators for *PTE* and η when true $\eta = 1$. Lower (upper) refers to the 2.5th (97.5th) percentile of the posterior. Bayesian point estimates are posterior medians. For $N = 40$, the MLE's for *PTE* and η are 0.435 and 1.231 respectively. The values are 0.744 and 0.851 for $N = 200$.

$\eta = 1$								
	$N = 40$				$N = 200$			
	Prior 1	Prior2	Prior3	Prior 4	Prior 1	Prior2	Prior3	Prior 4
<i>PTE</i>	0.483	0.671	0.297	0.461	0.685	0.751	0.536	0.727
Lower	0.118	0.345	0.057	0.037	0.400	0.505	0.251	0.410
Upper	0.846	0.671	0.632	0.918	0.906	0.934	0.769	0.967
η	1.065	0.975	1.149	1.076	0.859	0.812	0.909	0.849
Lower	-0.272	-0.332	-0.178	-0.258	0.345	0.291	0.394	0.34
Upper	2.372	2.318	2.533	2.416	1.365	1.316	1.408	1.359
<i>p</i> -value	0.0975				0.0014			

In contrast to Table 18, interval estimates were narrower. The priors had considerable impact on the posterior for *PTE* in the small sample case. Frequentist estimators for *PTE* were markedly inflated from 0.195 to 0.435. Moreover, the MLE and posterior median *PTE* using conditional uniform prior structures were in agreement. In the small sample case, the interval estimates of η for all three prior structures were rather wider in contrast to those in the larger sample case.

In this larger sample case, the posterior for PTE was not as influenced by the prior structure. The Bayesian point and interval estimates for both sample cases included the true value of η .

Figure 18 compares the posterior kernel density plots for PTE and η using prior1 with two different sample sizes (40 and 200).

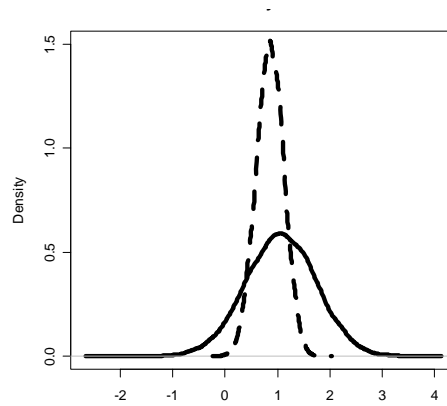


Figure 18. Posterior densities for η with samples size 40 and 200 using prior 1. The true value is $\eta = 1$. Here the solid line represents posteriors with $N = 40$ while the dashed line is for $N = 200$.

4.3.3 Scenario 3 ($\eta = 1.5$)

Table 20 displays the frequentist estimators and posterior median estimators for PTE and the interaction effect, η . The frequentist hypothesis testing p -value for each data set is also shown. The hypothesis of no interaction would be rejected in either case. For the case of $N = 40$, the MLE's for PTE and η are 0.453 and 2.335, respectively. The values are 0.947 and 1.389 in the case of $N = 200$.

Table 20 indicates that a large interaction effect ($\eta = 1.5$) inflates the value of PTE estimators; compare to Table 18. The Bayesian point and interval estimates for the

small sample case overestimate η . The true value of η is included in the interval estimates for all four prior structures and both sample sizes.

Figure 19 compares the posterior kernel density plots for η utilizing prior 1 under $N = 40$ and $N = 200$. In this case the Bayesian analysis agrees with the frequentist conclusion that interaction is present.

Table 20. Estimators for PTE and η when true $\eta = 1.5$. Lower (upper) refers to the 2.5th (97.5th) percentile of the posterior. Bayesian point estimates are posterior medians. For $N = 40$, the MLE's for PTE and η are 0.453 and 2.335 respectively.
The values are 0.947 and 1.389 for $N = 200$.

$\eta = 1.5$								
	$N = 40$				$N = 200$			
	Prior 1	Prior 2	Prior 3	Prior 4	Prior 1	Prior 2	Prior 3	Prior 4
PTE	0.500	0.690	0.304	0.475	0.844	0.868	0.711	0.903
Lower	0.121	0.355	0.056	0.038	0.633	0.685	0.485	0.680
Upper	0.867	0.934	0.645	0.942	0.971	0.977	0.875	0.995
η	2.057	1.962	2.099	2.084	1.425	1.395	1.483	1.398
Lower	0.862	0.801	0.916	0.906	0.899	0.886	0.962	0.892
Upper	3.295	3.203	3.354	3.32	1.92	1.895	1.985	1.885
p -value	0.0008				< 0.0001			

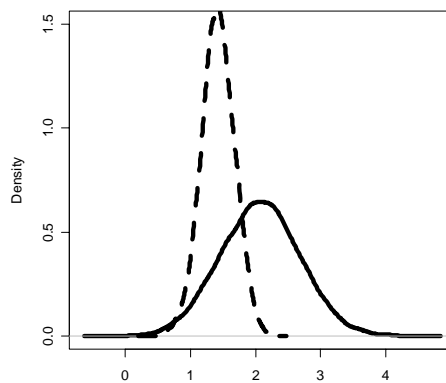


Figure 19. Posterior densities of η for samples with size 40 and 200 when true $\eta = 1.5$. Here the solid line represents posteriors with $N = 40$ while the dashed line is for $N = 200$.

4.4 Estimating PTE When Small Interaction is Present

As we have noted, significant interaction is an important concern in evaluating surrogate endpoints because the effect of the latter on the response becomes confounded with the treatment. For this reason, *PTE* is meaningless in the presence of significant interaction. However, if interaction is present but relatively small, can *PTE* still be used to evaluate surrogates?

Let us return to the model discussed in Section 4.1 where we defined *PTE* as a function of S , repeated here for convenience:

$$\begin{aligned} PTE(S) &= 1 - \frac{\beta'_\gamma + \eta S}{\beta} \\ &= 1 - \frac{\beta'_\gamma}{\beta} - \frac{\eta}{\beta} S \\ &= 1 - \frac{\beta'_\gamma}{\beta} - h(S). \end{aligned}$$

If $\beta'_\gamma \approx \beta_\gamma$ and $h(S)$ are small enough, we can ignore interaction in the model. But, how small should they be for practical estimation of *PTE*? Let $l(S)$ and $u(S)$ be the lower and upper bounds, respectively, of a 95% credible set on $h(S)$. Consider the following conditions:

- (1) Model (4.3) is at least approximately true;
- (2) $\beta'_\gamma \approx \beta_\gamma$;
- (3) $0 \in [l(S), u(S)]$ with $u(S) - l(S)$ small for all S .

Under these conditions, we have $PTE(S) \approx PTE$ for all S . Unfortunately, these conditions are fairly restrictive, as the following example illustrates.

In model (4.3) suppose T and S are defined as in Section 4.2. Set $\mu_\gamma = 0$, $\beta'_\gamma = 0.36$, $\gamma' = 0.7$, and $\eta = 0.2$. We generated $n = 40$ observations for this model, using the Poisson-normal assumptions of Section 4.2. Note that we did nothing to guarantee that $h(S)$ would be uniformly small, per the discussion in Section 4.1. To fit (4.1), (4.2) and (4.3) with this data we used WinBUGS, as in the previous examples. Parallel chains beginning at dispersed starting values were computed with 10,000 burn-in iterations and 30,000 iterations used for the posterior computations.

The posterior densities for β_γ and β'_γ are depicted in Figure 20. Their means (standard deviations) were 0.6246 (0.2961) and 0.6001 (0.2845), respectively. Clearly, it is reasonable to assume that $\beta_\gamma \approx \beta'_\gamma$.

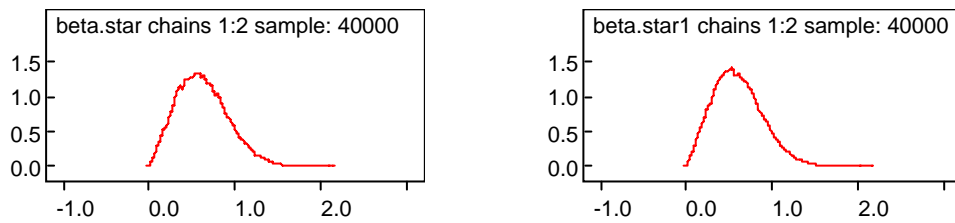


Figure 20. Posterior densities for β_γ and β'_γ .

Posterior densities for β and η are shown in Figure 21. Their means (standard deviations) were 0.4899 (0.696) and 1.625 (0.3639), respectively.

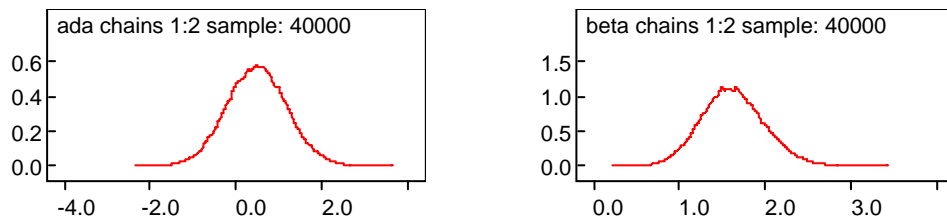


Figure 21. Posterior densities for β and η .

In this case, however, the quantity $h(S)$ is not sufficiently well behaved, as the plot in Figure 22 illustrates. That is, while condition (2) holds, condition (3) evidently fails, the median interval width being 1.056. (An analysis of this data set using other methods can be found in Section 4.3.1.).

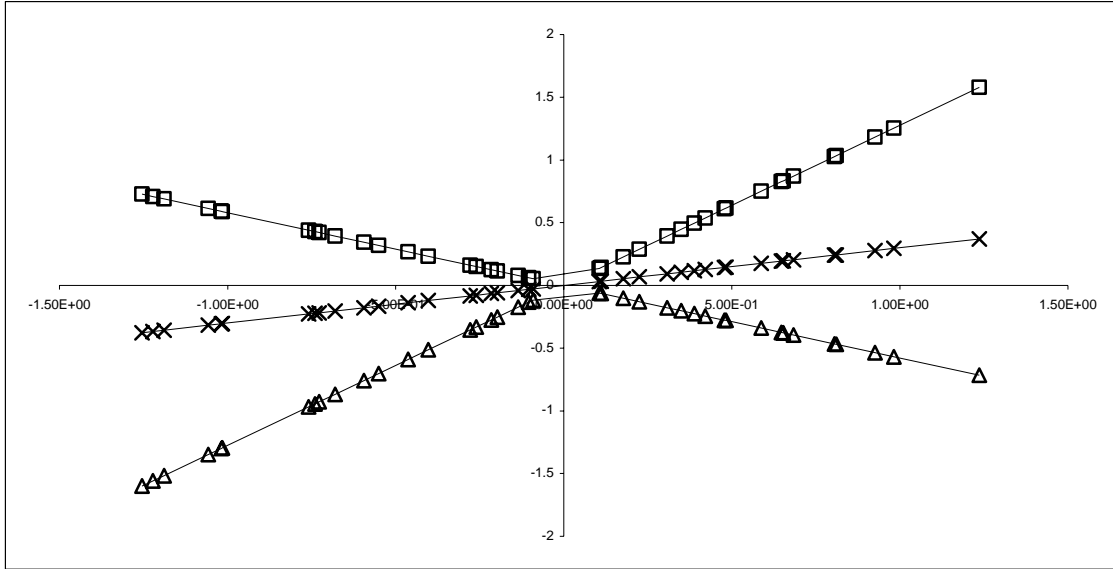


Figure 22. Median and 95% credible set values for $h(S)$. The abscissa is S . Median, upper, and lower bounds are indicated with \times , \square , and Δ respectively.

4.5 Joint Distribution of PTE and η

4.5.1 Introduction

Cowels (2002) did not specify a Bayesian method to assess the level of interaction presented in the data. In our fully Bayesian approach we prefer to model the uncertainty about interaction along with other parameters in the model. We can do this in several ways, but the most practical appears to be Bayesian model selection methods. Before we consider these, however, let us turn to the very special case in which there is expert consensus about what constitutes practically insignificant interaction.

Suppose expert opinion or previous data can establish what constitutes a practically insignificant interaction effect. That is, suppose it is known that $|\eta| < \delta$ is practically insignificant for some small $\delta > 0$. Suppose further that it is required that PTE be larger than a specified value, say, c . Then we can consider the joint posterior probability that $|\eta| < \delta$ and $PTE > c$. In practice, it will turn out that η and PTE are approximately independent, so that the joint posterior probability is the product of the marginal posterior probabilities; that is,

$$P(|\eta| < \delta \text{ and } PTE > c | \mathbf{E}, \mathbf{T}, \mathbf{S}) \approx P(|\eta| < \delta | \mathbf{E}, \mathbf{T}, \mathbf{S}) P(PTE > c | \mathbf{E}, \mathbf{T}, \mathbf{S}) \quad (4.9)$$

“Good” surrogate endpoints should have “large” values for both $P(|\eta| < \delta | \mathbf{E}, \mathbf{T}, \mathbf{S})$ and $P(PTE > c | \mathbf{E}, \mathbf{T}, \mathbf{S})$. Assuming approximate posterior independence between PTE and η , as is reasonable, one can represent this criterion graphically as in Figure 23.

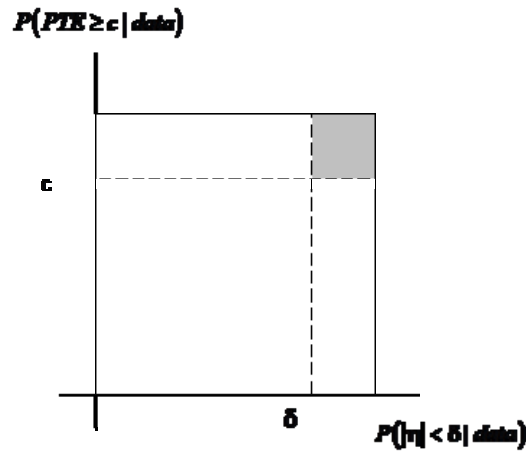


Figure 23. Marginal probabilities for PTE and $|\eta|$.

4.5.2 An Example

As an illustration of this approach, suppose $c = 0.7$ and $\delta = 1$. Using SAS we simulated $N = 200$ triples, (E_i, S_i, T_i) , using the Poisson-normal model described in Section 4.2 with $\beta'_\gamma = 0.1$, $\gamma' = 0.7$, and $\eta = 0.05$.

As in the last section, we placed the diffuse normal distribution, $N(0, 100,000)$ on each of μ , μ_γ , μ'_γ , and η . We used the conditional priors

$$\beta \sim \text{beta}_{[0,B]}(2, 2),$$

$$\beta_\gamma | \beta \sim \text{beta}_{[0,\beta]}(2, 5)$$

and

$$\beta'_\gamma | \beta \sim \text{beta}_{[0,\beta]}(2, 5).$$

We again chose $B = 5$. For this example, only one prior structure was used. The consequent induced prior on PTE is $\text{beta}_{[0,1]}(5, 2)$.

As before, MCMC methods were used to compute posterior quantities of interest. Convergence diagnostics indicated poor mixing so we increased the length of each chain from 10,000 to 50,000 after discarding 1,000 burn-in iterations. We retained every 10 sample in the chains.

Figure 24 displays the posterior distribution for PTE and η . The posterior marginal probability for $PTE > 0.7$ is 0.987 and the posterior marginal probability for $|\eta| < 1$ is 0.7054. The joint posterior probability that $PTE > 0.7$ and $|\eta| < 1$ is 0.698. Note that the product of the marginal probabilities equals the joint probability.

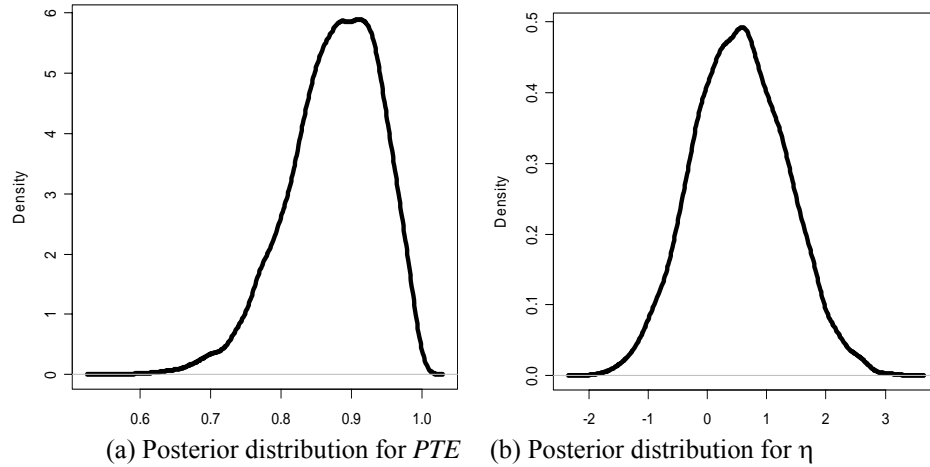


Figure 24. Posterior distributions for PTE and η .

4.6 Bayesian Model Comparisons

It would be unrealistic to count on routinely being able to specify what constitutes a practically insignificant interaction effect. How should one determine if interaction is to be retained in the model? Bayesian model comparison techniques are ideally suited for this purpose. In this section, we will discuss the use of the Bayesian information criterion (BIC) and the deviance information criterion (DIC) in the process of model selection.

4.6.1 Overview of Model Comparison

Consider a model indexed by a parameter vector $p \times 1$ parameter vector $\boldsymbol{\theta}$. Let $\hat{\boldsymbol{\theta}}$ denote the point estimates of $\boldsymbol{\theta}$ obtained using a suitable $n \times 1$ sample vector \mathbf{y} with likelihood l . BIC is defined as

$$BIC = -2\log l(\hat{\boldsymbol{\theta}} | \mathbf{y}) + 2p \log(n) \quad (4.9)$$

BIC can be interpreted as the discrepancy between data and model using a point estimate for $\boldsymbol{\theta}$. As such, smaller values of BIC indicated “better” models. Note that this criteria penalizes the number of parameters through the term $2p \log(n)$. BIC is considered as an

analogue to likelihood ratio test. BIC can be utilized to approximate Bayes factor via relation $BIC_1 - BIC_2 \approx -2\log(\text{Bayes Factor}_{12})$.

DIC can be viewed as a Bayesian analogue to the Akaike's information criteria (AIC). For each model, one obtains the posterior distribution of the log likelihood evaluated at the observed data and compare across models (Gelfand and Ghosh, 2000). That is, define

$$D(\boldsymbol{\theta}) = -2\log l(\boldsymbol{\theta} | \mathbf{y}) + 2\log h(\mathbf{y})$$

where, again, l is the likelihood and h is a standardizing function of the data alone. The latter has no affect on model selection. The fit of the model is measured by the expected value of the deviance with respect to the posterior distribution, $\mathcal{E}_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})]$. Model complexity is measured by the effective number of parameters, p_D , often defined as

$$p_D = \mathcal{E}_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] - D(\mathcal{E}_{\boldsymbol{\theta}|\mathbf{y}}[\boldsymbol{\theta}]).$$

Then we define

$$DIC = \mathcal{E}_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] + p_D. \quad (4.10)$$

Smaller values of DIC indicate “better” models. For more detail see Carlin and Louis (2000). Note that p_D serves as a penalty on the number of parameters in the model. DIC have been applied in a variety of research areas, such as pharmacokinetic modeling (Rhman *et al.*, 1999) and spline models with Bernoulli responses (Biller and Fahrmeir, 2001).

4.6.2 Bayesian Model Comparison Methods

In this section we compare the “null model” (2.1),

$$\mathcal{E}(E|T, S) = \mu_\gamma + \beta_\gamma T + \gamma S$$

with the “full model” (4.1),

$$\mathcal{E}(E | T, S) = \mu'_\gamma + \beta'_\gamma T + \gamma' S + \eta TS$$

In our example, the hypotheses are $H_0: \eta = 0$ versus $H_1: \eta \neq 0$. For this comparison, we used the data sets in Section 4.3.1 ($\eta = 0.2$) and Section 4.3.3 ($\eta = 1.5$). As before, we used WinBUGS to calculate posterior quantities of interest. To investigate robustness, we used three prior structures, originally defined in Section 4.3: prior 1 is left-skewed, prior 2 is symmetric, and prior 3 is right-skewed.

The likelihood ratio, LR , and AIC are frequentist model comparison statistics. They are defined as:

$$LR = -2 \left\{ \log \left[-l(\hat{\theta}_{null} | \mathbf{y}) \right] - \log \left[-l(\hat{\theta}_{full} | \mathbf{y}) \right] \right\} \quad (4.11)$$

and

$$AIC = -2 \log \left[-l(\hat{\theta} | \mathbf{y}) \right] + 2p \quad (4.12)$$

where $\hat{\theta}_{null}$ and $\hat{\theta}_{full}$ are MLE's under the null and full models, respectively, and $\hat{\theta}$ is the MLE for the null or full model.

Tables 21 and 22 contain values of DIC , AIC , BIC and the likelihood ratio for two models (4.1) and (2.1) for the case of $\eta = 0.2$. As can be seen in Table 21, as the sample size increases, all Bayesian model comparison statistics increased since they are functions of n . With moderate sample size ($n = 200$), the differences between null model and full model were slightly larger than those for the smaller sample size. There is little to choose between the models with respect to the different prior structures.

Table 22 contains details of the comparison when $\eta = 1.5$. The large interaction effect ($\eta = 1.5$) tended to inflate the model comparison statistics. Consequently, the absolute differences of statistics between two models increased as well.

For example, with prior 1 and $n = 40$, $DIC_{full} - DIC_{null} = -9.59$, which favors the full model. One would be ill-advised to use *PTE* as a measure of surrogate validity in such a case. The frequentist model comparison statistics (likelihood ratio and *AIC*) indicate the similar results.

Table 21. Model comparison statistics for scenario with $\eta = 0.2$. The WinBUGS quantites are $Dbar = -2 \log l(\theta | y)$ and $Dhat = 2 \log h(y)$.

$N = 40, AIC_{full} - AIC_{null} = 1.54, LR = 0.46$								
		Dbar	Dhat	p_D	DIC	$DIC_{full} - DIC_{null}$	BIC	$BIC_{full} - BIC_{null}$
Prior 1	Null Model	98.10	95.64	2.46	100.56	0.89	117.77	6.58
	Full Model	98.14	94.84	3.31	101.45		124.35	(0.037)*
Prior 2	Null Model	98.71	96.38	2.33	101.04	1.27	118.51	6.84
	Full Model	99.08	95.84	3.24	102.31		125.35	(0.037)
Prior 3	Null Model	97.80	95.49	2.31	100.11	0.80	117.62	6.45
	Full Model	97.73	94.56	3.17	100.91		124.07	(0.040)
Prior 4	Null Model	98.19	95.61	2.58	100.77	0.80	117.74	6.54
	Full Model	98.17	94.77	3.40	101.57		124.28	(0.038)
$N = 200, AIC_{full} - AIC_{null} = 1.9, LR = 0.1$								
		Dbar	Dhat	p_D	DIC	$DIC_{full} - DIC_{null}$	BIC	$BIC_{full} - BIC_{null}$
Prior 1	Null Model	511.18	508.48	2.70	513.87	1.91	540.27	10.49
	Full Model	512.08	508.37	3.70	515.78		550.76	(0.005)
Prior 2	Null Model	511.41	508.78	2.63	514.04	1.78	540.57	10.43
	Full Model	512.22	508.62	3.60	515.82		551.00	(0.005)
Prior 3	Null Model	511.36	508.87	2.49	513.85	1.98	540.66	10.55
	Full Model	512.32	508.82	3.50	515.83		551.21	(0.005)
Prior 4	Null Model	511.39	508.48	2.91	514.31	1.83	540.27	10.49
	Full Model	512.25	508.37	3.88	516.14		550.76	(0.005)

Note: Numbers in the parenthesis indicates the approximated Bayes factor.

4.6.3 Conclusions

In general, traditional hypothesis testing about interaction is problematic. We cannot assume that interaction effect is absent solely based on a p -value. As in any Bayesian analysis, we prefer to model uncertainty about unknowns. Therefore, we recommend examination of features of the joint posterior distribution of PTE and interaction in studying the validity of a surrogate endpoint. If bounds cannot be established for defining “practically insignificant” interaction, we recommend Bayesian model selection be used to determine the need for interaction in the model rather than hypothesis testing.

Table 22. Model comparison statistics for scenario with $\eta = 1.5$. The WinBUGS quantites are $Dbar = -2 \log l(\theta | y)$ and $Dhat = 2 \log h(y)$.

$N = 40, AIC_{full} - AIC_{null} = -9.45, \text{likelihood ratio} = 11.45$								
		Dbar	Dhat	p_D	DIC	$\frac{DIC_{full}}{-DIC_{null}}$	BIC	$\frac{BIC_{full}}{-BIC_{null}}$
Prior 1	Null Model	121.09	118.70	2.39	123.48	-10.99	140.83	-5.26
	Full Model	109.28	106.06	3.22	112.49		135.57	
Prior 2	Null Model	121.26	118.96	2.29	123.55	-10.32	141.09	-4.62
	Full Model	110.09	106.96	3.14	113.23		136.47	
Prior 3	Null Model	121.20	118.91	2.29	123.49	-11.49	141.04	-5.77
	Full Model	108.88	105.76	3.12	112.00		135.27	
Prior 4	Null Model	121.22	118.70	2.52	123.74	-11.17	140.83	-5.34
	Full Model	109.28	105.98	3.30	112.57		135.49	
$N = 200, AIC_{full} - AIC_{null} = -23.5, LR = 25.5$								
		Dbar	Dhat	p_D	DIC	$\frac{DIC_{full}}{-DIC_{null}}$	BIC	$\frac{BIC_{full}}{-BIC_{null}}$
Prior 1	Null Model	614.01	612.00	2.01	616.02	-23.57	643.79	-14.89
	Full Model	589.48	586.51	2.97	592.45		628.90	
Prior 2	Null Model	614.01	612.01	2.00	616.08	-23.62	643.80	-14.86
	Full Model	589.49	586.55	2.94	592.46		628.94	
Prior 3	Null Model	613.96	611.99	1.97	615.93	-23.57	643.78	-14.92
	Full Model	589.41	586.47	2.95	592.36		628.86	
Prior 4	Null Model	614.81	612.39	2.41	617.22	-24.72	644.18	-16.48
	Full Model	588.90	585.31	3.60	592.50		627.70	

Note: Numbers in the parenthesis indicates the approximated Bayes factor.

CHAPTER FIVE

Relative Effect

Although *PTE* has remained in use for the validation of surrogate endpoints, it has been severely criticized in the literature. Burzykowski, Molenberghs and Buyes (2005) have questioned the utility of *PTE* in clinical trials, noting that “with large clinical trials, the denominator of the *PTE* (the effect of treatment on the true end point – β) will be estimated with little precision, for otherwise, the need for a surrogate endpoint would no longer exist”. An alternative approach for evaluating surrogate endpoints was proposed by Buyse and Molenberghs (1998). Their method involves the relative effect (*RE*) and the association (ρ) between surrogate endpoints and true endpoints after adjustment for treatment.

In Section 5.1 we introduce *RE* and ρ . Inference for *RE* and ρ is discussed in Section 5.2. Section 5.3 we provide an overview of power priors, developed by Ibrahim and Chen (2000). The purpose of Section 5.4 is to illustrate the Bayesian analysis of *RE*. Three prior structures are used in that section: diffuse normal, informative, and power priors. The latter two prior structures are based on historical data.

5.1 Introduction

Buyse and Molenberghs (1998) studied frequentist estimation methods for *RE* and ρ for standard normal endpoints. We summarize their development here. As in previous chapters, let S denote the surrogate endpoint, T a binary treatment indicator, and E the true endpoint. For standard normal endpoints, we assume:

$$S_i = \mu' + \alpha T_i + \varepsilon_{S_i} \quad (5.1)$$

and

$$E_i = \mu + \beta T + \varepsilon_{E_i} \quad (5.2)$$

where

$$\begin{pmatrix} \varepsilon_S \\ \varepsilon_E \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right). \quad (5.3)$$

RE is defined as β/α and ρ is the correlation between S and E . A joint model for S and E given T is used to estimate RE and ρ simultaneously. A large sample size is recommended to obtain precise estimation for RE .

RE will be equal to 1 if the effects of treatment on the true endpoint are the same as those on the surrogate endpoint. In that case, it is considered a perfect surrogate endpoint at the population level. That is, regardless of any treatment effect, the true endpoint is mainly determined by the surrogate endpoint. If there is a perfect linear relationship between surrogate and treatment, ρ will be equal to 1. Then S is deemed a perfect surrogate endpoint at the individual level. An ideal surrogate endpoint should have S and E identical up to a deterministic transformation. That is, the closer RE and ρ are to 1, the better the surrogate endpoint.

For standard normal endpoints, we can model the relationship between the true endpoint, surrogate endpoint, and treatment as

$$E_i = \mu_\gamma + \beta_\gamma T + \gamma S + \varepsilon_{TS_j}$$

where $\beta_\gamma = \beta - \rho\alpha$, $\gamma = \rho$, and $Var(\varepsilon_{TS_j}) = 1 - \rho$. In this normal theory case, Buyse and

Molenberghs (1998) derived a simple relationship between PTE , RE , and ρ :

$$PTE = \rho \frac{\alpha}{\beta} = \rho \frac{1}{RE}. \quad (5.4)$$

5.2 Inference for RE

5.2.1 Frequentist Inference for RE and ρ

Buyse and Molenberghs (1998) developed maximum likelihood estimators for RE and ρ for binary and standard normal endpoints. The latter uses models (5.1), (5.2), and (5.3). Fieller's theorem or the delta method is needed to estimate confidence limits for RE (Henson, 1975). See Section 1.2 for details of this development. Buyse (2000) augured that RE can be typically estimated with greater precision than PTE . The effect, β , of the treatment on the true endpoint is the denominator for calculating PTE . However, this parameter is frequently poorly estimated when the confidence limits of PTE are acceptably narrow. In contrast, the denominator of RE , α (the effect of the treatment on the surrogate endpoint) is precise enough for the confidence limits of RE to be informative.

For non-standard normal endpoints, we use (5.1) and (5.2) and assume the covariance structure

$$\begin{pmatrix} \varepsilon_S \\ \varepsilon_E \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{SS}^2 & \sigma_{SE} \\ \sigma_{SE} & \sigma_{EE}^2 \end{pmatrix} \right). \quad (5.5)$$

Let $\mu_S = \mu' + \alpha \mathcal{E}(T)$ and $\mu_E = \mu + \beta \mathcal{E}(T)$. The conditional bivariate normal density for S and $E|T$ is

$$\begin{aligned} f(S, E | T) = & \left(2\pi\sigma_{SS}\sigma_{EE}\sqrt{1-\rho^2} \right)^{-1} \\ & \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{S-\mu_S}{\sigma_{SS}} \right)^2 - \frac{2\rho(S-\mu_S)(E-\mu_E)}{\sigma_{SS}\sigma_{EE}} + \left(\frac{E-\mu_E}{\sigma_{EE}} \right)^2 \right] \right\} \end{aligned} \quad (5.6)$$

where $-\infty < S < \infty$, $-\infty < T < \infty$, $\sigma_{SS} > 0$, $\sigma_{EE} > 0$, and $-1 < \rho < 1$.

Since the magnitude of the endpoint variances will influence the magnitude of the RE , Cowels (2004) used the standardized RE , defined as

$$RE = \frac{\beta / \sigma_{EE}}{\alpha / \sigma_{SS}} \quad (5.7)$$

where α and β are estimated for model (5.1) and (5.2), and σ_{EE}^2 , σ_{SS}^2 are the variance components for ε_E and ε_S , respectively, as in (5.4). The adjusted association ρ is defined to be $\rho = \sigma_{SE} / \sqrt{\sigma_{SS}^2 \sigma_{EE}^2}$, where σ_{SE} is the covariance between E and S .

Molenberghs, Geys and Buyse (2002) extended the estimation for RE and ρ to the model with mixed discrete and continuous endpoints. The calculation of RE and ρ for multiple clinical trials was also studied. A mixed model was built to quantify the surrogacy at the trial level and individual level.

Burzykowski, Molenberghs and Buyse (2005) also noted two potential difficulties in using RE and ρ . First, RE might not be practical since confidence intervals for it are often too wide to be informative. Furthermore, using RE requires the strong assumption that the relationship between the treatment effects on the surrogate and the true endpoints is multiplicative. A multiplicative relationship may not obtain if RE changes with the strength of the association between treatment and the outcomes themselves.

5.2.2 Bayesian Inference for RE and ρ for Longitudinal Data

Cowels (2004) developed Bayesian inferential methods for RE and ρ . She extended the approaches of Buyse and Molenberghs (1998) to a longitudinal surrogate endpoint S and time-to-event data with a censored true endpoint E . She worked with

three types of models: log normal, Weibull regression, and joint normal-proportional hazards (PH).

Cowels (2004) used a flat prior structure on the coefficients in the linear model. An informative Whishart prior was assumed for the precision matrix. Cowels (2004) did not consider Bayesian inference for RE and ρ for normal endpoints. In the following section, we focus on Bayesian analysis for RE and ρ given normal endpoints.

5.2.3 Bayesian Inference for RE and ρ for Normal Endpoints

Denote the joint bivariate likelihood of model (5.6) by $l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|S, E, T)$, where the variance-covariance matrix for S and E is given by (5.5). Let the prior distributions for the coefficients in model (5.1), (5.2) be denoted by $p(\boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\mu, \mu', \alpha, \beta)$, and the prior distribution for the variance-covariance matrix, $\boldsymbol{\Sigma}$, by $p(\boldsymbol{\Sigma})$. Then the joint posterior distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|S, E, T) \propto p(\boldsymbol{\beta})p(\boldsymbol{\Sigma})l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|S, E, T). \quad (5.8)$$

As in previous chapters, we used MCMC methods to draw samples from the joint posteriors. At convergence, the MCMC simulation produces a sequence of values of $\boldsymbol{\beta}$ from the joint posterior (5.8). Values of (α, β) from this sequence can then be used to construct the posterior distribution of RE . Values of σ_{SS}, σ_{SE} , and σ_{EE} from the chain after convergence can be used to construct the posterior of ρ .

We used three choices for $p(\boldsymbol{\beta})$: diffuse normal, informative normal, and a power prior. The latter two are based on the historical data. In the next section, we introduce the background for power priors.

5.3 Power Priors

When historical data is available, Bayes' theorem can be used to fit a posterior distribution which subsequently can be used as a prior for new data. This is usually accomplished by beginning with a relatively non-informative prior. If the historical likelihood is especially “peaked”, the resulting posterior may be overly precise. That is, the resulting posterior may overwhelm the likelihood from the new data. Ibrahim and Chen (2000) have developed a new class of prior distributions for use in this scenario. Their power priors attenuate the influence of the historical data on the current likelihood.

Several researchers have applied power priors to various problems:

- Chen, Manatunga and Williams (1998) utilized the power prior for heritability estimates from human twin data;
- Chen, Ibrahim, and Yiannoutsos (1999) used power priors in variable selection for logistic regression;
- Chen, Ibrahim, Shao and Weiss (1998) used power priors for the class of generalized linear mixed models and focused on variable selection;
- Chen, Ibrahim and Sinha (1999) developed power priors for various types of models for survival data.

Let D denote the data from the “current” study and let $l(\boldsymbol{\theta}|D)$ be the corresponding, “current” likelihood, where $\boldsymbol{\theta}$ is a vector of parameters. Suppose D_0 is the data from the “historical” study and denote by $l(\boldsymbol{\theta}|D_0)$ the “historical” likelihood function. Let $\pi_0(\boldsymbol{\theta}|\mathbf{c}_0)$ represent the “initial” prior distribution for $\boldsymbol{\theta}$ to be used with D_0 , indexed by the parameter vector \mathbf{c}_0 . Here, “initial” indicates prior structures to be used in

conjunction with the historical data. Then the power prior distribution of θ for the current study is defined as

$$\pi(\theta|D_0, a_0) \propto l(\theta|D_0)^{a_0} \pi_0(\theta|\mathbf{c}_0), \quad (5.9)$$

where a_0 is a parameter that weights the “historical” likelihood relative to the “current” likelihood. Here $a_0 = 1$ is equivalent to specifying a prior for θ that is proportional to the “historical” likelihood and $a_0 = 0$ corresponds to prior specification without incorporation of historical data. Specifying an appropriate value for a_0 in (5.9) may be problematic. The beta distribution has proven particularly useful as a prior for a_0 ; let $\pi(a_0|a, b)$ denote a beta prior for a_0 with parameters a and b . Then the joint power prior distribution for (θ, a_0) becomes

$$\pi(\theta, a_0|D_0) \propto l(\theta|D_0)^{a_0} \pi_0(\theta|\mathbf{c}_0) \pi(a_0|a, b). \quad (5.10)$$

Using (5.10), the posterior distribution for (θ, a_0) is

$$\pi(\theta|D, D_0, a_0) \propto l(\theta|D_0)^{a_0} \pi_0(\theta|\mathbf{c}_0) \pi(a_0|a, b) l(\theta|D) \quad (5.11)$$

Thus, the posterior distribution based on a power prior is equivalent to the posterior based on the product of the two likelihood functions $l(\theta|D) l(\theta|D_0)^{a_0}$.

5.4 An Example

In this section, we illustrate how to use historical data to construct the posterior for RE and ρ . We use diffuse and informative normal priors, as well as a power prior.

5.4.1 Data Generation

We simulated two data sets, one “historical” and one “current”, with sample sizes 100 and 40, respectively. In each data set, half of patients are in the treatment group

with $T = 1$ and half in the control group with $T = 0$. For simplicity, we generated bivariate standard normal endpoints.

For the “current” study data, the surrogate endpoints, S , and true endpoints, E , were generated from a bivariate normal distribution,

$$\begin{pmatrix} S \\ E \end{pmatrix} \Big|_T \sim N_2 \left(\begin{bmatrix} \mu_S T \\ \mu_T T \end{bmatrix}, \begin{bmatrix} \sigma_{SS_c}^2 & \sigma_{SE_c} \\ \sigma_{SE_c} & \sigma_{EE_c}^2 \end{bmatrix} \right) \quad (5.12)$$

Specially, we chose $\mu_S = 1.3$, $\mu_T = 1$ and

$$\Sigma_c \equiv \begin{bmatrix} \sigma_{SS_c}^2 & \sigma_{SE_c} \\ \sigma_{SE_c} & \sigma_{EE_c}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}.$$

With $\sigma_{SE_c} = 0.3$ we have $RE = 0.77$. Using the generated 40 observations, the MLE for RE is 1.07 with estimated variance-covariance matrix

$$\begin{bmatrix} 0.98 & 0.607 \\ 0.607 & 1.19 \end{bmatrix}$$

and estimated correlation 0.563. Note that the estimated correlation is larger than the true value of ρ (0.3), and MLE for RE is not close to the true value (0.77).

For the “historical” study, the surrogate endpoints, S , and true endpoints, E , were generated from (5.12) with $\mu_S = 1.5$, $\mu_T = 1.2$, and

$$\Sigma_h \equiv \begin{bmatrix} \sigma_{SS_h}^2 & \sigma_{SE_h} \\ \sigma_{SE_h} & \sigma_{EE_h}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Here $\sigma_{SE_c} = 0.5$ so we have $RE = 0.8$. Using the generated 100 observations, the MLE for

RE is 0.797 with estimated variance-covariance matrix $\begin{bmatrix} 1.74 & 0.937 \\ 0.937 & 1.12 \end{bmatrix}$

and estimated correlation 0.67, which is larger than the true value of ρ (0.5), and MLE for RE is close to the true value (0.77).

5.4.2 Prior Structure

Power Prior

The power prior for $\beta = (\mu, \mu', \alpha, \beta)$ can be specified based on the “historical” likelihood:

$$\pi(\beta, a_0 | D_0) \propto l(\beta | D_0)^{a_0} \pi_0(\beta | \mathbf{c}_0) \pi(a_0 | a, b) \quad (5.13)$$

The joint posterior distribution for β and the precision matrix $\mathbf{T} \equiv \Sigma_h^{-1}$ is

$$\pi(\beta, \mathbf{T} | a_0, D_0, D) \propto l(\beta | D_0)^{a_0} \pi_0(\beta, | \mathbf{c}_0) \pi(a_0 | a, b) l(\beta, \mathbf{T} | D) \pi(\mathbf{T}). \quad (5.14)$$

To investigate the effect of the power parameter a_0 on the posterior in (5.14), we used three beta distributions with parameters (2, 90), (2, 2), and (90, 2). For comparison purposes we also included the fixed values $a_0 = 0$ and $a_0 = 1$.

We chose initial priors (π_0) for β to be a diffuse normal, with zero mean vector, conditional on \mathbf{T} . A Wishart(\mathbf{R}, k) prior was specified for \mathbf{T} , the precision matrix for the bivariate normal endpoints. The Wishart distribution we used has density

$$f(\mathbf{T}) = |\mathbf{R}|^{k/2} |\mathbf{T}|^{(k-p-1)/2} \exp[-0.5 \text{tr}(\mathbf{RT})] \quad (5.15)$$

where $p = 2$, \mathbf{T} is symmetric and positive definite, and \mathbf{R} is a scale matrix.

To represent vague prior knowledge, we chose the degrees of freedom, k , to be as small as possible (i.e. 3). We selected $\mathbf{R} = \begin{bmatrix} 200 & 0 \\ 0 & 0.2 \end{bmatrix}$

This prior is summarized in Table 23.

Table 23. Initial prior distribution in power prior distribution.

Initial Prior distribution	
$\left. \begin{array}{c} \mu \\ \beta \\ \mu' \\ \alpha \end{array} \right\}$	$N(0, 100,000)$
\mathbf{T}	$\text{Wishart}\left(\left[\begin{array}{cc} 200 & 0 \\ 0 & 0.2 \end{array}\right], 3\right)$

The posterior (5.14) is not of closed form. We used MCMC methods to obtain a sample of posterior values. After convergence, pairs (α, β) can be drawn from this sample and the posterior for RE constructed from them.

Informative Prior

To obtain a (non-power) prior distribution using the historical data, we set an initial diffuse normal prior structure for β , along with a diffuse Wishart for the precision matrix. The posterior is then obtained in the usual way. Independent normal priors to be used with the current data are then placed on the components of β by matching moments with the posterior obtained using the historical data, as was done in Section 3.3.

The precision matrix, \mathbf{T} , was given a Wishart prior. To select the parameters for this Wishart prior, we proceeded as follows (see Bernardo and Smith, 1993). Based on the posterior means for precision matrix and variance-covariance matrix, computed using the historical data, we can obtain estimates of the parameters in the Wishart distribution (5.15). Specifically, we have

$$\mathcal{E}(\mathbf{T}) = (k/2)\mathbf{R}^{-1}$$

and

$$\mathcal{E}(\mathbf{T}^{-1}) = \left(\frac{k - p - 1}{2} \right)^{-1} \mathbf{R}$$

where $p = 2$ and \mathbf{T}^{-1} is the variance-covariance matrix. Using the posterior mean for the \mathbf{T} and \mathbf{T}^{-1} , we derived the \mathbf{R} matrix for the Wishart distribution, which is

$$\begin{bmatrix} 137.4 & 27.8 \\ 27.8 & 62.65 \end{bmatrix}$$

and $k = 97$.

Table 24 lists the informative prior structures that will be used for the “current” data. The WinBUGS model is listed in Appendix E.4 and R code to call WinBUGS is included in Appendix D.5.

Table 24. Informative normal prior distribution.

Prior distribution	
μ	$N(0.212, 0.146)$
β	$N(1.077, 0.207)$
μ'	$N(0.176, 0.165)$
α	$N(1.344, 0.228)$
\mathbf{T}	$\text{Wishart}\left(\begin{bmatrix} 137.4 & 27.8 \\ 27.8 & 62.65 \end{bmatrix}, 97\right)$

Diffuse Normal Prior

To compare with the results with informative prior and power prior, diffuse normal prior structures were also used. In this scenario, we discarded the information from the “historical” data, using instead the diffuse prior $N(0, 100,000)$. The precision matrix is assumed to follow a diffuse Wishart distribution as described above. This prior

is summarized below in Table 25. The WinBUGS model is listed in Appendix E.4 and R code to call WinBUGS is included in Appendix D.5.

Table 25. Diffuse normal prior distribution.

Prior distribution	
$\left. \begin{array}{c} \mu \\ \beta \\ \mu' \\ \alpha \end{array} \right\}$	$N(0, 100,000)$
\mathbf{T}	$\text{Wishart}\left(\left[\begin{array}{cc} 200 & 0 \\ 0 & 0.2 \end{array}\right], 3\right)$

5.4.3 Simulation Results for Power Prior

We sampled from the posterior (5.11) using MCMC methods implemented in WinBUGS. The WinBUGS program is included in Appendix E.3. The R program to call WinBUGS is in the Appendix D.4. We used two chains with dispersed initial values and checked for convergence using Gelman-Rubin statistics and autocorrelation plots. There was within-sequence correlation during draws. To counter this we thinned the chains, keeping every 5th simulation draw from each sequence and discarding the rest starting with 1,000 burn-in iterations and 20,000 sample iterations. Subsequent convergence diagnostics were satisfactory. Table 26 displays the posterior inferences for RE and p .

The posterior estimations for RE combine the information from the “current” likelihood and “historical” likelihood, the latter via the power prior. The joint likelihood from the “historical” data and “current” data dominates RE posterior calculations. The 95% credible sets for RE include 1. Therefore, at the population level, the effect of treatment T on the true endpoint E relative to that of treatment effect on surrogate endpoints for this example is large.

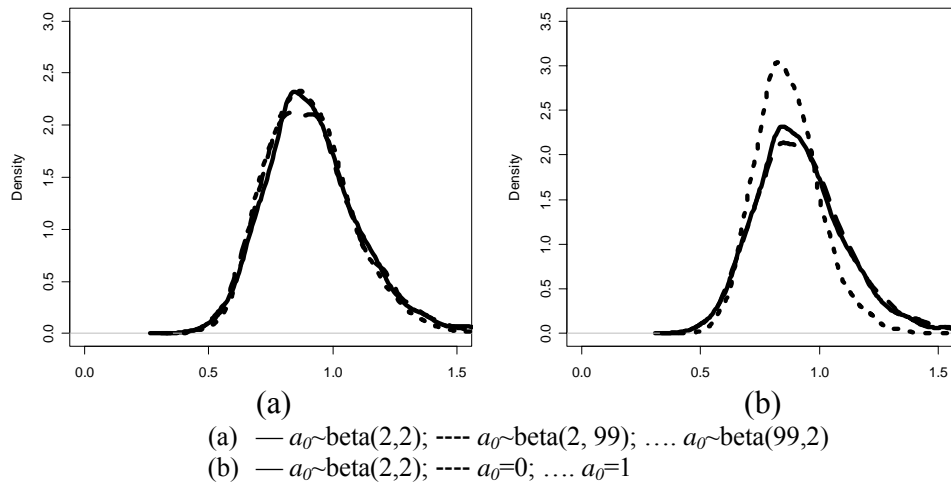
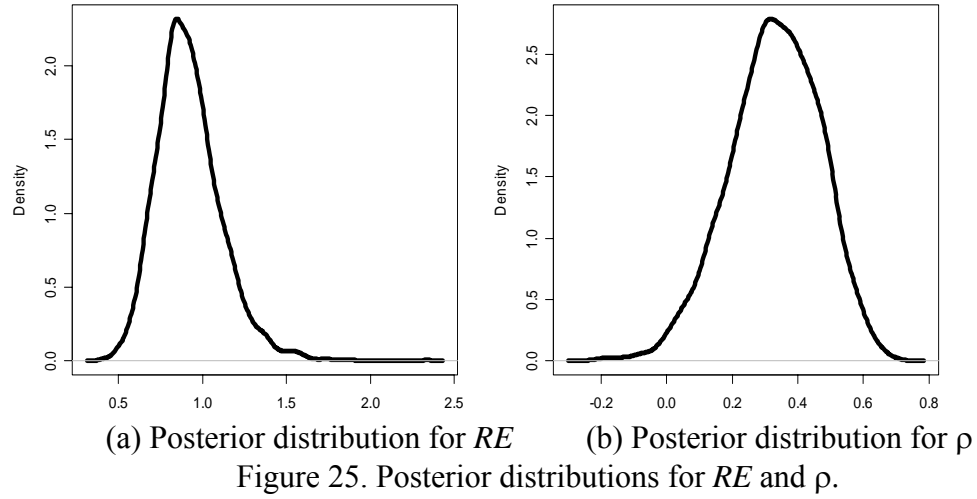
Table 26. Posterior and MLE estimations for RE , ρ , α , β and Σ . True values for “current” data and “historical” data are also listed. The power prior in Table 23 was used in this analysis.

Parameter	“Current”		“historical”		Mean (95% credible set)
	True	MLE	True	MLE	
RE	0.77	1.07	0.8	0.797	0.917 (0.594, 1.362)
ρ	0.3	0.563	0.5	0.67	0.328 (0.04, 0.580)
α	1.3	0.944	1.5	1.346	1.201 (0.840, 1.562)
β	1	1.01	1.2	1.074	1.077 (0.761, 1.386)
$Var(S)$	1	0.98	1	1.74	1.252 (0.806, 1.933)
$Cov(S,E)$	0.3	0.607	0.5	0.937	0.374 (0.041, 0.804)
$Var(E)$	1	1.19	1	1.12	1.016 (0.649, 1.566)

The posterior estimate for ρ is close the true value for the “current” data (-0.3). Note that the prior information for ρ is not dependent on the “historical” data since we used a diffuse Wishart prior for the precision matrix. Consequently, posterior estimates are close the true values for the “current” data.

Figure 25 exhibits the posterior distributions for RE and ρ with different a_0 . The initial prior structure for power a_0 is assumed to be $\text{beta}(2, 2)$.

We also investigated the effect of the power parameter, a_0 , on the posterior. In Figure 26 we have graphed the posterior densities for RE corresponding to different choices for a_0 . Figure 26(a) shows the posterior distributions for RE when we set a_0 to be fixed. Recall from Section 5.3 that $a_0 = 1$ is equivalent to using a scaled “historical” likelihood for the prior on β and $a_0 = 0$ is corresponds to a prior on β with no incorporation of historical data.



We also considered three beta priors on a_0 : $\text{beta}(2, 2)$, $\text{beta}(2, 90)$, and $\text{beta}(90, 2)$. Figure 26(b) shows the posterior distributions for RE under these three prior structures for a_0 .

Since our data were generated from similar distributions (albeit with very different sample sizes), the “historical” and “current” likelihood functions are quite similar. As a result, in our examples, changes in a_0 did not dramatically affect posterior estimations of RE .

5.4.4 Comparison of Posteriors under Power, Informative, and Diffuse Priors

In addition to the posterior comparisons made above, we also compared the power prior with an informative Wishart component. This prior structure is summarized in Table 27 below. Table 28 contains posterior summaries for RE , α and β under the four prior structures.

Table 27. Power Prior with informative Wishart structure.

Initial Prior distribution	
μ β μ' α	$\left. \begin{array}{l} \mu \\ \beta \\ \mu' \\ \alpha \end{array} \right\} N(0, 100,000)$
T	Wishart $\left(\begin{bmatrix} 137.4 & 27.8 \\ 27.8 & 62.65 \end{bmatrix}, 97\right)$

Table 28. Posterior mean and 95% credible sets for RE , α , β and ρ . The true value of RE is 0.77. Column letters are for ease of reference in the text.

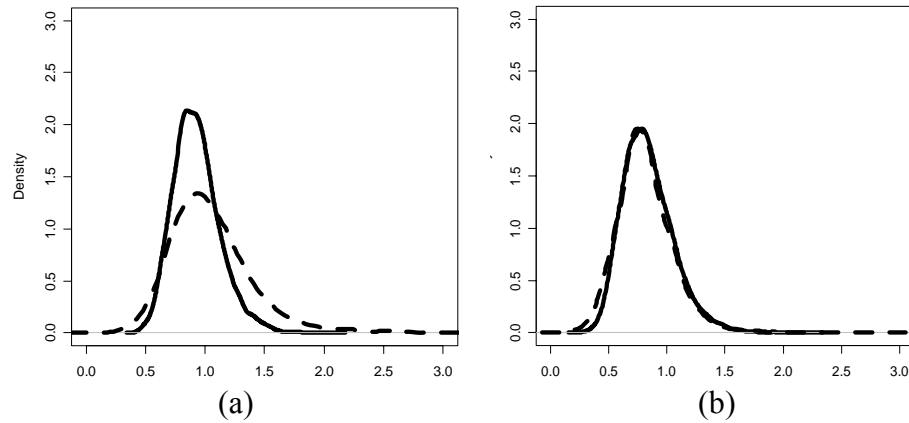
	Power prior				Informative prior (Table 24)	Diffuse normal prior (Table 25)
	$a_0=0$ (Table 23)	$a_0=1$ (Table 23)	$a_0 \sim \text{beta}(2, 2)$ (Table 27)	$a_0=1$ (Table 23)		
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>
RE	0.924 (0.6, 1.37)	0.872 (0.637, 1.186)	0.939 (0.613, 1.37)	0.880 (0.644, 1.176)	1.045 (0.486, 1.882)	0.921 (0.61, 1.38)
α	1.197 (0.843, 1.565)	1.253 (0.970, 1.528)	1.176 (0.840, 1.51)	1.238 (0.979, 1.508)	1.037 (0.562, 1.532)	1.245 (0.814, 1.589)
β	1.081 (0.774, 1.393)	1.079 (0.843, 1.312)	1.083 (0.760, 1.401)	1.078 (0.841, 1.314)	1.028 (0.519, 1.538)	1.009 (0.71, 1.319)
ρ	0.332 (0.050, 0.582)	0.330 (0.040, 0.583)	0.325 (0.171, 0.468)	0.324 (0.166, 0.471)	0.324 (0.165, 0.469)	0.331 (0.035, 0.591)

Not surprisingly, use of a diffuse normal prior resulted in wider 95% credible sets.

The power prior with $a_0 = 1$ and an informative Wishart (Column *D* in Table 28) showed improvement in the credible set for ρ , as expected. Similarly, the informative prior with

$a_0 \sim \text{beta}(2, 2)$ (Column *C* in Table 28) exhibited decreased width for the credible set on ρ compared to the informative and diffuse priors. The posterior mean overestimates RE for all of these prior choices. Likelihoods dominate the posterior for RE . Recall that the data we generated has a higher correlation between surrogate endpoints and true endpoints. This might be the reason for the overestimated RE .

We have also graphed some of the posteriors described in Table 28. Figure 27(a) compares the posterior distribution for RE using an informative normal prior (Column *E* in Table 28) with the posterior under a power prior when $a_0 = 1$ (Column *D*). Figure 27(b) makes the same comparison but between Column *A* and Column *F*. Again, a diffuse Wishart prior was assumed for the precision matrix in these two cases.



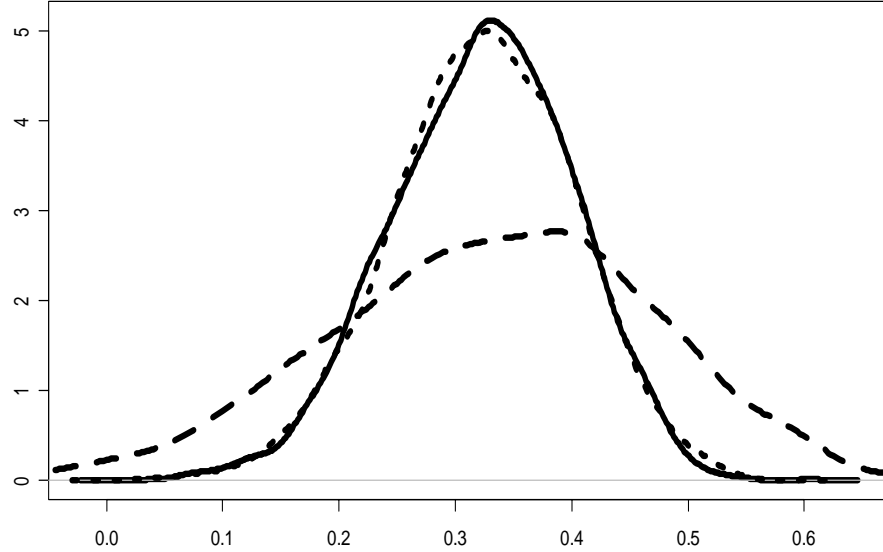
(a) —: power prior with $a_0=1$; --- informative normal prior

(b) —: power prior with $a_0=0$; ---- diffuse normal prior

Figure 27. Posterior distributions for RE with different prior structures. In Figure (a) we compared column B and E. In Figure (b) we have column A and F.

In Figure 28 we compare the posterior distribution for RE using an informative normal prior (Column *E* in Table 28), a diffuse normal prior (Column *F* in Table 28) with the posterior under a power prior when $a_0 = 1$, this is the case where an informative

Wishart prior was assumed for the precision matrix (Column D in Table 28). Similar results are concluded from the comparisons between Column C , E , and F .



Note: solid line: informative prior; dash line: power prior with diffuse Wishart prior; dot line: power prior with informative Wishart prior

Figure 28. Posterior distributions for ρ with different prior structures. In this figure, we have used the structures in Columns C , E and F of Table 28.

5.5 Power Prior for Bayesian Estimation for PTE

This Section focuses on a simple example illustrating a Bayesian analysis for *PTE* with normal endpoints using a power prior approach. We used the same data sets as in Section 5.4.

5.5.1 Bayesian Estimation for PTE with a Power Prior

For normal endpoints, we assumed that the true endpoints, treatment and surrogate endpoint have the following relationship:

$$E_i|T_i, S_i = \mu_\gamma + \beta_\gamma T + \gamma S + \varepsilon_{TS_i} \quad (5.16)$$

and

$$E_i|T_i = \mu + \beta T + \varepsilon_{T_i} \quad (5.17)$$

where ε_{TS_i} and ε_{T_i} are independently normally distributed.

Again, the proportion of treatment effects explained by surrogate endpoints was defined as $PTE = 1 - \beta_\gamma / \beta$.

We used the power prior for $\beta \equiv (\mu, \mu_\gamma, \beta_\gamma, \beta, \gamma)$ defined by

$$\pi(\beta, a_0 | D_0) \propto l(\beta | D_0)^{a_0} \pi_0(\beta | c_0) \pi(a_0 | a, b) \quad (5.18)$$

where $l(\beta | D_0)$ denotes the joint likelihood function for models (5.16) and (5.17) and, again, D_0 denoted the “historical” data D_0 .

The joint posterior distribution for β is

$$\pi(\beta | a_0, D_0, D) \propto l(\beta | D_0)^{a_0} \pi_0(\beta | c_0) \pi(a_0 | a, b) l(\beta | D) \quad (5.19)$$

where, as before, D denotes the current data set. We used two initial prior structures for β and β_γ : diffuse normal prior structure and conditional beta prior structures. Independent diffuse normal priors are placed for coefficients μ , μ_γ , and γ . The power parameter, a_0 , is assumed to follow a beta(2, 2) distribution. Diffuse gamma distributions were assigned to precision components τ_1 and τ_2 . The resulting joint prior is summarized in Table 29.

Table 29. Initial prior distributions in power prior method.
(a) β and β_γ follow diffuse normal prior

Initial Prior distribution	
μ β μ_γ β_γ	$\left. \begin{array}{l} \\ \\ \\ \end{array} \right\} N(0, 100,000)$
τ_1 τ_2	$\left. \begin{array}{l} \\ \end{array} \right\} \text{gamma}(0.01, 0.01)$

Table 29. Initial prior distributions in power prior method (continued).
 (b) β and β_γ follow conditional beta prior

Initial Prior distribution	
μ	$N(0, 100,000)$
μ_γ	$N(0, 100,000)$
β	$\text{beta}_{[0, 5]}(2, 2)$
β_γ	$\text{beta}_{[0, \beta]}(2, 2)$
τ_1	$\text{gamma}(0.001, 0.001)$
τ_2	

5.5.2 Posterior Results under the Power Prior

As before we used MCMC methods to sample from the joint posterior in (5.18), implemented in WinBUGS. The program is in Appendix E.2 and R code to call WinBUGS model is listed in Appendix D.4. We used 20,000 sample iterations with 1,000 burn-in iterations, starting with dispersed initial values in two chains. After thinning the chains (every 5th iteration was retained) Gelman-Rubin statistics and autocorrelation plots indicated no problems with convergence.

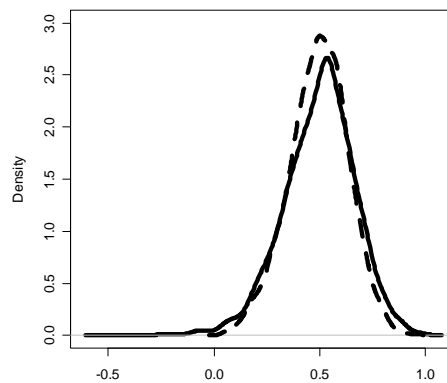


Figure 29. Posterior distribution for *PTE* with power prior. Here the solid line indicates use of a diffuse normal as the initial prior and the dash line indicates use of an initial beta prior.

The MLE for PTE is 0.35 for “current” data and 0.56 for “historical” data. Figure 29 displays the posterior distribution for PTE under the power prior. Two prior structures are assumed in this figure: diffuse normal prior and conditional beta priors. See details in table 29.

With an initial diffuse normal prior structure, the posterior mean for PTE is 0.498 with 95% credible set (0.138, 0.804). In contrast, using a conditional beta initial prior, the posterior mean for PTE is 0.487 and the 95% credible set is (0.205, 0.761).

CHAPTER SIX

Summary and Future Research

To save time and reduce the size and cost of clinical trials, surrogate endpoints are frequently measured instead of true endpoints. The proportion of the treatment effect explained by surrogate endpoints (*PTE*) is a widely used validation criteria. Frequentist and Bayesian methods have been developed to facilitate such validation. Unfortunately, with both methods proposed to date estimates of *PTE* can be outside the unit interval. To preclude this problem, we proposed a new model using conditional beta priors, applying it both Poisson distributed true endpoints and normally distributed surrogate endpoints. Furthermore, we derived the induced prior on *PTE* thereby allowing direct comparison of prior and posterior distributions. A Cauchy distribution centered at 1 resulted when diffuse normal priors are used, which proved quite unsuitable. We also introduced the use of power priors for *PTE*.

We proposed a fully Bayesian approach to account for uncertainty about interaction. Inference for both *PTE* and the interaction effect were considered. Also regarding interaction, we considered Bayesian model comparison techniques.

We performed Bayesian analysis of the relative effect (*RE*) and the association between surrogate endpoints and true endpoints (ρ) as an alternative to *PTE*. We used power priors to develop the posteriors for *RE* and ρ as well.

Our research considered surrogate endpoint evaluation for a single clinical trial. It is clearly of interest to study validation criterion for multiple trials. In the future, we hope to extend our investigation to include meta-analysis.

Our focus was on single surrogate endpoints. Cowles (2002) provided Bayesian framework for multiple covariates with binary endpoints. Her work would be a good starting point from which to apply our ideas to the case of multiple covariates, particularly in a generalized linear model context.

Finally, missing value and misclassification problems are very important in clinical trials generally and therefore in surrogate endpoints evaluation. It is of interest to extend our investigation to situations involving missing values or misclassification.

APPENDICES

APPENDIX A

Convergence Issues in MCMC Simulations

There are convergence difficulties in the iterative simulation inferences. Iterative simulation draws with within-sequence correlation are generally less precise than the same number of independent draws. Such correlation can cause inefficiencies in simulations.

Brooks and Gelman (1998) and Best, Cowles and Vines (1997) summarized the convergence diagnostics. One is the Gelman & Rubin diagnostics (1992) which propose a general approach to monitoring convergence of MCMC output in which two or more parallel chains are run with over-dispersed initial values. Values substantially above 1 indicate lack of convergence.

The other diagnostic is autocorrelation plot. High autocorrelations within chains indicate slow convergence. That is iterative simulation draws with within-sequence correlation.

For instance, Figure A.1 displays the autocorrelation for draws from posterior *PTE* distribution in one chain for one data set in section 3.3. Obviously, there is high within-sequence correlation during draws. About 20 sequentially draw of *PTE* are correlated with each other. The statistics estimate from these draws can not represent the true distribution of posterior *PTE*.

There are three approaches to handle special convergence problems. First, we can manipulate multiple simulation sequences with starting points dispersed. Two

parallel simulation chains with different starting points are used for every simulation study in this chapter. Second, we monitor the prior structures and increase the precision in the priors from parameters. We decreased the variance for diffuse normal prior structure. If the above two ways can not effectively solve the convergence issue, we will use the third method, the most important one, that we thin the sequences by keeping every k th simulation draw from each sequence and discarding the rest. Best, *et al.* (1997) suggest increasing the thinning interval to say, every 5th or 10th iteration, before calculating summary statistics and density estimates, in order to achieve a less highly correlated sample.

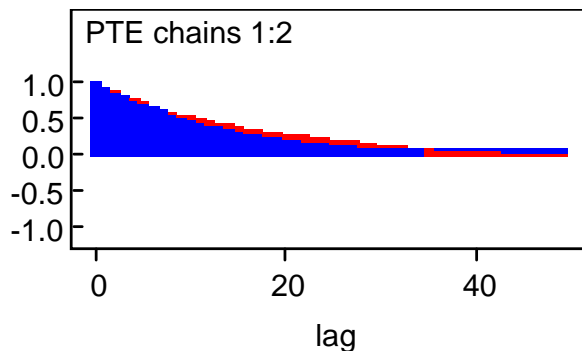


Figure A.1 Autocorrelation for one chain of posterior *PTE* distribution before thin the sequences.

For example, we will be able to make the simulation sequence longer, say, from 10000 to 20000, and thin the sequences by keeping every 10th simulation draw from each sequence and discarding the rest.

Figure A.2 shows the autocorrelation for draws from posterior *PTE* distribution after thinning the sequences by keeping every 10th draw. Apparently, after increasing

the length of the chain and the thinner sequences, the with-sequence correlation of draws noticeably decreased.

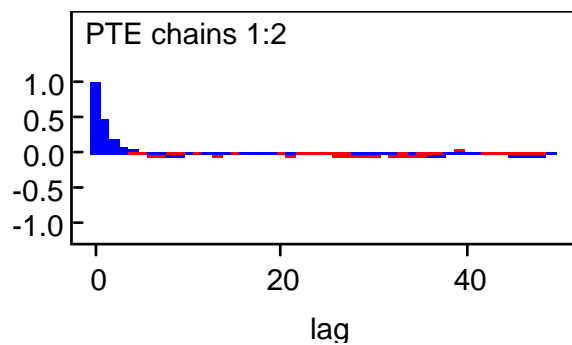


Figure A.2 Autocorrelation for one chain of posterior *PTE* distribution after thin the sequences.

APPENDIX B

SAS IML Program to Generate Data

B.1. SAS Programs to Generate Data with Poisson Distributed Endpoints and Normally

Distributed Surrogate Endpoints

```

proc iml;
  n=250;
  E1=j(n,1,1); mu1=j(n,1,1);
  E2=j(n,1,1); mu2=j(n,1,1);
  S1=j(n,1,1); S2=j(n,1,1);
  T1=j(n,1,0.5); T2=j(n,1,-0.5);
  seed1=1; seed2=2;
  do i=1 to n;
    S1[i,1]=1+rannor(seed1);
    S2[i,1]=-1+rannor(seed2);
    mu1[i,1]=exp(0.375*T1[i,1]+0.875*S1[i,1]);
    mu2[i,1]=exp(0.375*T2[i,1]+0.875*S2[i,1]);
    E1[i,1]=ranpoi(seed1,mu1[i,1]);
    E2[i,1]=ranpoi(seed2, mu2[i,1]);
  end;
  T=T1//T2;
  S=S1//S2;
  E=E1//E2;
  variables=T||S||E;
  create scenario from variables[colname={T S E}];
  append from variables;
  quit;

```

B.2. SAS Programs to Generate Data for Bivariate Normally Distributed Endpoints

```

proc iml;
  n=20; m=50;
  S=j(2*n,1,1);
  E=j(2*n,1,1);
  z1=j(2*n,1,0); z2=j(2*n,1,0); z3=j(2*n,1,0);
  z4=j(2*m,1,0); z5=j(2*m,1,0); z6=j(2*m,1,0);
  Sh=j(2*m,1,0); Eh=j(2*m,1,0);
  T1=j(n,1,1); T2=j(n,1,0);
  T=j(2*n, 1,0); rho=0.3;
  T=T1//T2;
  Th1=j(m,1,1); Th2=j(m,1,0);
  Th=j(2*m, 1,0); rho_h=0.5;
  Th=Th1//Th2;
  seed1=12345; seed2=4567;seed3=62341;
  seed4=137967; seed5=3595;seed6=78141;
  do i=1 to 2*n;
    * generate 'current data';

```

```

z1[i,1]=rannor(seed4);  z2[i,1]=rannor(seed5);
z3[i,1]=rannor(seed6);
S[i,1]=1.3*T[i,1]+1*(sqrt(1-rho)*z1[i,1]+sqrt(rho)*z3[i,1]);
E[i,1]=1*T[i,1]+1*(sqrt(1-rho)*z2[i,1]+rho*z1[i,1]);
end;
do i=1 to 2*m;
  * generate 'historical data';
  z4[i,1]=rannor(seed4);
  z5[i,1]=rannor(seed5);
  z6[i,1]=rannor(seed6);
  Sh[i,1]=1.5*Th[i,1]+1*(sqrt(1-
rho_h)*z4[i,1]+sqrt(rho_h)*z6[i,1]);
  Eh[i,1]=1.2*Th[i,1]+1*(sqrt(1-
rho_h)*z5[i,1]+sqrt(rho_h)*z6[i,1]);
end;
variables1=T||S||E;
create scenario1 from variables1[colname={T S E}];
append from variables1;
variables2=Th||Sh||Eh;
create scenario2 from variables2[colname={Th Sh Eh}];
append from variables2;

quit;

```

APPENDIX C

SAS Procedure to Perform Frequentist Analysis

C.1. SAS GENMOD Procedure to Analysis Data When True Endpoints Follow Poisson

Distribution

* To test the significance of interaction effect;

```
proc genmod data=scenario;
    model E = T S T*S / dist    = poisson
                        link     = log;
run;
```

* To build Poisson regression between E and T, S;

```
proc genmod data=scenario;
    model E = T S / dist    = poisson
                        link  = log;
run;
```

* To build Poisson regression between E and T;

```
proc genmod data=scenario;
    model E = T / dist    = poisson
                link     = log;
run;
```

C.2 SAS REG Procedure to Analysis Data for Bivariate Normal Endpoints

* To build regression between E and T, S;

```
proc reg data=scenario1;
    model E = T S ;
run;
```

* To build regression between E and T;

```
proc reg data=scenario1;
    model E = T ;
run;
```

* To build regression between S and T;

```
proc reg data=scenario1;
    model S= T ;
run;
```

APPENDIX D

D.1. R program to Calculate MLE's for *PTE* in Chapter 3

```
##### Getting frequentist estimators for 500 data sets#####
iter_freq<-function() {
  N<-2*n
  freq<-matrix(rep(0,4*k),nrow=k,ncol=4)
  t<-0
  ##### start iteration #####
  for (i in 1: k) {
    E1<-true[,i]
    E2<-true[,i]
    S<-surrogate[,i]
    T<-treat[,i]

    ##### Frequentist estimate of beta, beta.star and PTE
    glm_ETS<-glm(E1~S+T+S*T, family=poisson()) ### find how many data has p-value that is less
                                                than 0.05
    glm_ET<-glm(E2~T,family=poisson())
    glm_ETS1<-glm(E1~S+T, family=poisson())
    freq[i,1]<-summary(glm_ETS1)$coefficients[3,1] ### frequentist estimator of beta.star
    freq[i,2]<-summary(glm_ET)$coefficients[2,1]   ### frequentist estimator of beta
    freq[i,3]<-1-freq[i,1]/freq[i,2]              ### frequentist estimator of PTE
    freq[i,4]<-summary(glm_ETS)$coefficients[4,4] ### p-value of interaction effect
    if (freq[i,3]<0 ) t<-t+1                      ### test how many significant p-value
  }
  return(t,freq)
}
```

D.2. R Program to WinBUGS to Get Bayesian Estimation for *PTE* with Conditional Beta

Priors

```
#####Getting Posterior Estimates for PTE Using beta priors #####
iter_beta<-function () {
  N<-2*n
  PTE.freq<-matrix(rep(0,k),nrow=k,ncol=1)
  beta1.freq<-matrix(rep(0,k),nrow=k,ncol=1)
  beta2.freq<-matrix(rep(0,k),nrow=k,ncol=1)
  CS<-matrix(rep(0,5*k),nrow=k,ncol=5)
  ##### start iteration #####
  for (i in 1: k) {

    E1<-true[,i]
    E2<-true[,i]
    S<-surrogate[,i]
    T<-treat[,i]
```

```
##### Call winbugs to calculate posterior distribution of PTE ###
data<-list("N","T","S","E1","E2","a","b","c","d","f","g")
inits1<-list(mu.star=0,gamma.star=0.1, b1=0.1, b2=0.5,mu=0.5, mu.S=0,tau.S=1)
inits2<-list(mu.star=0.1, gamma.star=0.3, b1=0.3, b2=0.7,mu=1,mu.S=0.5,tau.S=1)
inits<-list(inits1, inits2)
parameters<- c("mu.star","beta.star","gamma.star","mu","beta","PTE")

# poisson.bug contain the Winbugs program for bivariate binomial distribution
poisson.sim<-
bugs(data,inits,parameters,"poisson.bug",n.chains=2,n.thin=1,n.iter=sim,n.burnin=1000,
debug=FALSE)
summary<-poisson.sim$summary
CS[i,1]<-summary[6,1] ### posterior mean of PTE
CS[i,2]<-summary[6,2] ### posterior std dev. of PTE
CS[i,3]<-summary[6,3] ### posterior 2.5% of PTE
CS[i,4]<-summary[6,5] ### posterior median of PTE
CS[i,5]<-summary[6,7] ### posterior 97.5% of PTE
}
return (CS)
}
```

D.3. R Program to WinBUGS to Get Bayesian Estimation for *PTE* with Truncated or Diffuse Normal Priors

```
##### Iterations for truncated/ normal priors #####
iter_normal<-function () {
  N<-2*n
  CS<-matrix(rep(0,5*k),nrow=k,ncol=5)

##### start iteration #####
  for (i in 1: k) {
    E1<-true[,i]
    E2<-true[,i]
    S<-surrogate[,i]
    T<-treat[,i]
    data<-list("N","T","S","E1","E2")

    inits1<-list(mu.star=0,gamma.star=1, beta=0.5,beta.star=0.1, mu=0.3, mu.S=0, tau.S=1)
    inits2<-list(mu.star=0.1, gamma.star=0.5, beta=0.7,beta.star=0.5,mu=0.1,mu.S=0.1,tau.S=1)
    inits<-list(inits1, inits2)
    parameters<- c("mu.star","beta.star","gamma.star","mu","beta","mu.S","PTE")
    poisson_normal.sim<- bugs(data, inits, parameters,"Poisson_truncated.bug",
n.chains=2,n.thin=1, n.iter=sim,n.burnin=1000, debug=FALSE)
    summary<-poisson_normal.sim$summary
    CS[i,1]<-summary[1,1] ### posterior mean of PTE
    CS[i,2]<-summary[1,2] ### posterior std dev. of PTE
    CS[i,3]<-summary[1,3] ### posterior 2.5% of PTE
    CS[i,4]<-summary[1,5] ### posterior median of PTE
    CS[i,5]<-summary[1,7] ### posterior 97.5% of PTE
  }
  return (CS)
}
```

D.4 R-code to WinBUGS to Get Bayesian Estimations for *RE* with Power Prior

```
normal<-function() {
  data<-list("N","M","T","E","Eh","Sh","Th","R","a","b")
  inits1<-list(mu=0, mu.t=-0.375, beta=0.5,alpha=0.7,a0=1,tau1=1,
  tau2=0.6,tau=structure(.Data=c(1,0.5,0.5,1),.Dim=c(2,2)))
  inits2<-list(mu=0.1,mu.t=-0.5, beta=0.5, alpha=0.7,
  a0=0.5,tau2=0.2,tau1=0.2,tau=structure(.Data=c(1,0.8,0.8,1),.Dim=c(2,2)))
  inits<-list(inits1, inits2)
  parameters<- c("mu","mu.t","beta","alpha","RE","sigma","rho","tau")
  normal.sim<- bugs(data, inits, parameters,"normal.bug", n.chains=2,n.thin=5,
  n.iter=sim,n.burnin=1000, debug=TRUE)
  summary<-normal.sim$summary
  return (summary)
}
```

D.5. The R code to Call WinBUGS Program to Get Bayesian Estimations for *RE* with Informative or Diffuse Normal Prior

```
normal<-function() {
  data<-list("M","T","E","R")
  inits1<-list(mu=0, mu.t=-0.375, beta=0.5,
  alpha=0.7,tau=structure(.Data=c(1,0.5,0.5,1),.Dim=c(2,2)))
  inits2<-list(mu=0.1,mu.t=-0.5, beta=0.5, alpha=0.7,
  tau=structure(.Data=c(1,0.8,0.8,1),.Dim=c(2,2)))
  inits<-list(inits1, inits2)
  parameters<- c("mu","mu.t","beta","alpha","RE","tau","rho","sigma")
  normal.sim<- bugs(data, inits, parameters,"hist_normal_RE.bug", n.chains=2,n.thin=20,
  n.iter=sim,n.burnin=1000, debug=TRUE)
  summary<-normal.sim$summary
  return (summary)
}
```

D.6. R Program to Simulate the Induced Prior for *PTE* with Conditional Prior Method

```
prior<-function(a,b,c,d,f,g,sim) {
  ## initialized the variables
  b1<-matrix(rep(0,sim),nrow=sim,ncol=1)
  b2<-matrix(rep(0,sim),nrow=sim,ncol=1)
  beta<-matrix(rep(0,sim),nrow=sim,ncol=1)
  beta.star<-matrix(rep(0,sim),nrow=sim,ncol=1)
  PTE<-matrix(rep(0,sim),nrow=sim,ncol=1)
  for (i in 1:sim) {
    b1[i,1]<-rbeta(1,a,b)
    b2[i,1]<-rbeta(1,f,g)
    beta[i,1]<-b1[i,1]*(d-c)+c
    beta.star[i,1]<-b2[i,1]*(beta[i,1]-c)+c
  }
  PTE<-1-(beta.star/beta)
  return (PTE,beta,beta.star)
}
```

APPENDIX E

WinBUGS Programs

The following lines of code used in WinBUGS provide the basic structures for all models we implemented in WinBUGS. This basic structure contains the Bayesian estimations for *PTE*, *RE* and ρ with different prior structures.

E.1. WinBUGS Model for Poisson Distributed True Endpoint to Calculate *PTE*

Three prior structures for β and β_γ are constructed in this model structure: diffuse normal, truncated normal and conditional beta prior. If we would like to include interaction term in the model, simple adjustment for model structure is needed.

```
model
{
  for( i in 1 : N ) {
    # model E|T, S
    E1[i] ~ dpois(est1[i])
    # model E|T
    E2[i] ~ dpois(est2[i])
    S[i]~dnorm(mu.S, tau.S)
    log(est1[i]) <- mu.star + beta.star* T[i] +gamma.star * S[i]
    log(est2[i]) <- mu + beta * T[i]

  }
  # define diffuse priors
  mu.star~dnorm(0,0.000001)
  mu ~ dnorm(0,0.000001)
  gamma.star~dnorm(0,0.000001)
  mu.S~dnorm(0,0.000001)
  tau.S~dgamma(0.001, 0.001)

  # conditional beta priors  $\beta$  and  $\beta_\gamma$ 
  b1~dbeta(a,b)
  b2~dbeta(f,g)
  beta<-b1*(d-c)+c
  beta.star<-b2*(beta-c)+c
  # truncated normal priors for  $\beta$  and  $\beta_\gamma$ 
  # beta~djl.dnorm.trunc(0,0.0001,0,1000)
  # beta.star~djl.dnorm.trunc(0,0.0001, 0, beta)
  # diffuse normal priors for  $\beta$  and  $\beta_\gamma$ 
  # beta~dnorm(0, 0.00001)
```

```

#    beta.star~dnorm(0,0.00001)

PTE<-1-beta.star/beta
# If include interaction term, calculate joint/marginal distribution for PTE and  $\eta$ 
#Joint <- step(PTE - .7)*step(.3 - abs(ada))
#PTElarge <-step(PTE - .7)
#adasmall <-step(0.3-abs(ada) )
}

```

E.2 WinBUGS Models to Calculate *PTE* with Normal Endpoints using Power Prior. We used one-tricks twice to build power prior for *PTE*. Two initial prior structures are constructed in this program: diffuse normal and conditional beta prior.

```

model
# model for "current data"
{ for (i in 1:N) {
  E1[i]~dnorm(p1[i], tau1)
  E2[i]~dnorm(p2[i],tau2)
  p1[i]<-mu+beta*(T[i])
  p2[i]<-mu.star+beta.star*(T[i])+gamma.star*(S[i])
}
# ones trick
C <- 10000 # this just has to be large enough to ensure all p[i]'s < 1
# likelihood model for historical data
for (j in 1:M) {
  Eh1[j]~dnorm(phi[j],tau1)
  phi[j]<-mu+beta*(Th[j])
  Eh2[j]~dnorm(theta[j],tau2)
  theta[j]<-mu.star+beta.star*(Th[j])+gamma.star*(Sh[j])
  l1[j]<-(tau1/(2*3.1415926))*exp(-(Eh1[j]-phi[j])*(Eh1[j]-phi[j])*(tau1*tau1)/2)
  l2[j]<-(tau2/(2*3.1415926))*exp(-(Eh2[j]-theta[j])*(Eh2[j]-theta[j])*(tau2*tau2)/2)
# likelihood function for E|T,S
  ones1[j] <- 1
  pp[j] <- pow(l1[j],a0)/C
  ones1[j]~dbern(pp[j])
# likelihood function for E|T
  ones2[j] <- 1
  tt[j] <- pow(l2[j],a0)/C
  ones2[j]~dbern(tt[j])
}
# define the prior structure for the historical data and current data
mu~dnorm(0,0.00001)
mu.star~dnorm(0,0.00001)
gamma.star~dnorm(0, 0.00001)
# diffuse normal prior structure for and
# beta~dnorm(0, 0.00001)
# beta.star~dnorm(0,0.00001)
# Conditional beta prior structures for and
d~dbeta(2,2)
c~dbeta(2,7)
beta<-d*5
beta.star<-c*beta

```

```

# consider power  $a_0$  is a random variable
a0~dbeta(a,b)
# assign diffuse gamma distributions for precisions
tau1~dgamma(0.01, 0.01)
tau2~dgamma(0.01, 0.01)
PTE<-1-beta.star/beta
}

```

E.3 WinBUGS Models for Power Prior to Calculate *RE*. We used ones-trick twice to construct power prior for *RE*.

```

model
# model for "current data"
{ for (i in 1:N) {
  E[i, 1:2]~dmnorm(p[i,1:2], tau[1:2,1:2])
  p[i,1]<-mu+beta*(T[i])
  p[i,2]<-mu.t+alpha*(T[i])
}
}
# try ones trick
C <- 10000 # this just has to be large enough to ensure all p[i]'s < 1
# likelihood model for historical data
for (j in 1:M) {
  Eh[j]~dnorm(phi[j],tau1)
  phi[j]<-mu+beta*(Th[j])
  Sh[j]~dnorm(theta[j],tau2)
  theta[j]<-mu.t+alpha*(Th[j])
  l1[j]<-(sqrt(tau1)/(2*3.1415926))*exp(-(Eh[j]-phi[j])*(Eh[j]-phi[j])*tau1/2)
  l2[j]<-(sqrt(tau2)/(2*3.1415926))*exp(-(Sh[j]-theta[j])*(Sh[j]-theta[j])*tau2/2)
# likelihood function for E|T
  ones[j] <- 1
  pp[j] <- pow(l1[j]*l2[j],a0)/C
  ones[j]~dbern(pp[j])
}
# define the prior structure for the historical data and current data
mu~dnorm(0,0.01)
beta~dnorm(0, 0.01)
mu.t~dnorm(0, 0.01)
alpha~dnorm(0, 0.01)
a0~dbeta(a,b)
# diffuse gamma priors for precisions in historical data
tau1~dgamma(0.01, 0.01)
tau2~dgamma(0.01, 0.01)
# diffuse Wishart prior for precision matrix in current data
tau[1:2,1:2]~dwish(R[1:2,1:2],2)
# get variance-covariance matrix
t<-tau[1,1]*tau[2,2]-tau[2,1]*tau[1,2]
sigma[1,1]<-tau[2,2]/t
sigma[1,2]<-tau[1,2]/t
sigma[2,1]<-tau[2,1]/t
sigma[2,2]<-tau[1,1]/t
# get RE and rho
RE<-beta/alpha
rho<-sigma[1,2]/sqrt(sigma[1,1]*sigma[2,2]) }

```

E.4. WinBUGS Program to Model RE with Diffuse/Informative Normal Prior

```

model
# model for "current data"
{ for (i in 1:M) {
  E[i, 1:2]~dmnorm(p[i,1:2], tau[1:2,1:2])
  p[i,1]<-mu+beta*(T[i])
  p[i,2]<-mu.t+alpha*(T[i])
}

# define the prior structure for the historical data and current data
mu~dnorm(0.212,6.85)
beta~dnorm(1.077,4.83)
mu.t~dnorm(0.176,6.06)
alpha~dnorm(1.344, 4.386)
# mu~dnorm(0,0.01)
# beta~dnorm(0,0.01)
# mu.t~dnorm(0, 0.01)
# alpha~dnorm(0, 0.01)

tau[1:2,1:2]~dwish(R[1:2,1:2],94)

t<-tau[1,1]*tau[2,2]-tau[2,1]*tau[1,2]
sigma[1,1]<-tau[2,2]/t
sigma[1,2]<-tau[1,2]/t
sigma[2,1]<-tau[2,1]/t
sigma[2,2]<-tau[1,1]/t
#RE<-(beta/sqrt(sigma[1,1]))/(alpha/sqrt(sigma[2,2]))
RE<-beta/alpha
rho<-sigma[1,2]/sqrt(sigma[1,1]*sigma[2,2])
}

```

REFERENCES

- Akaike, H. (1983), "Information Theory and Model Selection", *Bulletin of the International Statistical Institute*, 50, 277-290.
- Alston, C., Kuhnert, P., Choy, S.L., McVinish, R. and Mengersen, K. (2005), "Bayesian Model Comparison: Review and Discussion", website: http://www.stat.auckland.ac.nz/~iase/publications/13/Alston-Kuhnert-Low_Choy-McVinish-Mengersen.pdf.
- Best, N.G., Cowles M.K. and Vines S.K. (1997), "CODA Convergence Diagnosis and Output Analysis Software for Gibbs sampling output", Version 0.4, Medical Research Council Biostatistics Unit, Cambridge, U.K.
- Brooks, S.P. and Gelman A. (1998), "Alternative Methods for Monitoring Convergence of Iterative Simulations", *Journal of Computational and Graphical Statistics*, 7, 434-455.
- Burzykowski, T., Molenberghs, G. and Buyse M. (2005), "The Evaluation of Surrogate Endpoints", Springer Science.
- Buyse M., and Molenberghs G., (1998), "Criteria for the Validation of Surrogate Endpoints in Randomized Experiments", *Biometrics*, 54, 014-1029.
- Buyes M., Molenberghs G., Burykowski T., Renard D. and Geys H. (2000), "Statistical Validation of Surrogate Endpoints: Problems and Proposals", *Drug Information Journal*, 34, 447-454.
- Carlin, B.P. and Louis, T.A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd Edition, Chapman and Hall, Boca Raton.
- Chaloner, K.M. and Duncan, G.T. (1983) "Assessment of beta prior distribution: PM elicitation." *The Statistician*, Vol. 32, 174-180.
- Chen, C., Wang, H. and Snapinn, S. (2003), "Proportion of Treatment Effect (PTE) Explained by A Surrogate Marker". *Statistics in Medicine*, 22, 3440-3459.
- Chen, M.-H., Ibrahim, J.G., Shao, Q.-M., and Weiss, R.E. (1999), "Prior Elicitation for Model Selection and Estimation in Generalized Linear Mixed Models", Technical report MS-06-99-02, Department of Mathematical Sciences, Worcester Polytechnic Institute.

- Chen, M.-H., Ibrahim, J.G., and Sinha, D. (1999), "A New Bayesian Model for Survival Data with A Surviving Fraction", *Journal of American Statistician Association*, 94, 909-919.
- Chen, M.-H., Ibrahim, J.G., and Yiannoutsos, C. (1999), "Prior Elicitation, Variable Selection and Bayesian Computation for Logistic Regression Models", *Journal of Royal Statistics Society, Series B*, 61, 223-242.
- Chen, M.-H., Manatunga, A.K. and Williams, C.J. (1998), "Heritability Estimates from Human Twin Data by Incorporating Historical Prior Information", *Biometric*, 54, 1348-1362.
- Cowles, Mary Kathryn. (2002). "Bayesian estimation of the proportion of treatment effect captured by a surrogate marker", *Statistics in Medicine* 21, 811-834.
- Cowels M. (2004), "Evaluating Surrogate Endpoints for Clinical Trials: A Bayesian Approach", Technique report, University of Iowa.
- Cummings SR, Black DM, Thompson DE, et al. (1998), "Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial", *Journal of American Medical Association*, 280, 2077-2082.
- Fleming, T. R. and DeMets, D. L. (1996), "Surrogate End Points in Clinical Trials: Are We Being Misled?", *Annals of Internal Medicine*, 125(7), 605-613.
- Freedman, L.S. (2001), "Confidence Intervals and Statistical Power of the 'Validation' Ratio for Surrogate or Intermediate Endpoints". *Journal of Statistical Planning and Inference*, 96, 143-153.
- Freedman, L.S., Graubard, B.I. and Schatzkin, A. (1992), "Statistical Validation of Intermediate Endpoints for Chronic Diseases", *Statistics in Medicine*, 11(2), 167-178.
- Gelman, A. and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences", *Statistical Science*, 7, 457-72.
- Gelfand, A.E. and Ghosh, M. (2000) "Generalized linear models: A Bayesian view," in *Generalized Linear Models*, edited by D.K. Dey, S.K. Ghosh, and B.K. Mallick, Dekker, New York, 3-21.
- Herson J. (1975), "Fieller's Theorem Versus the Delta Method for Significance Intervals for Ratios", *Journal of Statistical Computing and Simulation*, 3, 265-274.
- Ibrahim, J.G., and Chen, M.-H. (1998), "Prior Distributions and Bayesian Computation for Proportional Hazards Models", *Sankhya Series B*, 60, 48-64.

- Ibrahim J.G. and Chen M.-H. (2000), "Power Prior Distributions for Regression Models", *Statistical Science*, 15, 46-60.
- Li, Z., Meredith, M.P. and Hoseyni, M.S. (2001), "A Method to Assess the Proportion of Treatment Effect Explained by A Surrogate Endpoint", *Statistics in Medicine*, 20, 3175-3188.
- Lin, D.Y., Fleming, T.R. and DeGruttola, V. (1997), "Estimating the Proportion of Treatment Effect Explained by A Surrogate Marker". *Statistics in Medicine*, 16(3), 1515-1527.
- Lunn, D. (2004), "WBDev shared components", website: <http://www.winbugs-development.org.uk/>.
- Molenberghs, G., Burzykowski, T., Alonso, A. (2004), "A Perspective on Surrogate Endpoints in Controlled Clinical Trials". *Statistics in Medicine*, 13, 177-206.
- Molenberghs, G., Buyse, M., and Burzykowski, T. (2005), *The Evaluation of Surrogate Endpoints*, New York: Springer-Verlag, 65-82.
- Molenberghs, G., Geys, H. and Buyse, M (2002), "Evaluation of Surrogate Endpoints in Randomized Experiments with Mixed Discrete and Continuous Outcomes", *Statistics in Medicine*, 20, 3023-3038.
- Prentice, R.L. (1989), "Surrogate Endpoints in Clinical Trials: Definitions and Operational Criteria", *Statistics in Medicine*, 8, 431-440.
- Spiegelhalter, D., Best, N., Carlin, B., and Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit", *Journal of Royal Statistical Society, Series B*, 64, 583-639.
- Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1995), "BUGS Examples, Version 0.5, Volume 1", Medical Research Council Biostatistics Unit, Cambridge, U.K.
- Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003): *WinBUGS User Manual Version 1.4*, website: <http://www.mrc-su.cam.ac.uk/bugs/winbugs/manual14.pdf>.
- Sturtz, S., Ligges, U. and Gelman A. (2005), "R2WinBUGS: A Package for Running WinBUGS from R", *Journal of Statistical Software*, 12, 1-12.
- Wang Y. and Taylor (2002), "A Measure of the Proportion of Treatment Effect Explained by A Surrogate Marker". *Biometrics*, 5.