ABSTRACT

Contributions to the Theory and Practice of Prior Elicitation
in Biopharmaceutical Research

Somer Blair, Ph.D.

Mentors: John W. Seaman, Jr. and David J. Kahle

In this dissertation, we consider modeling problems in biopharmaceutical research, much of which is motivated by industry colleagues. Expert opinion is necessary in many applications of survival analysis, especially in exploratory and early phase research. We develop methods for eliciting informative priors using expert knowledge on observable survival time summaries in the proportional hazards model. In problems with small sample sizes and censoring, incorporating information from historical studies can enhance statistical inference. To this end, we present methods for selecting a critical parameter in the power prior using operational assessments of such choices, such as FDA guidance and prior effective sample size. We investigate the consequences of misspecified information in prior elicitations and create a mathematical framework and graphical guide with which to understand the effects. Finally, we investigate the effect of various non-informative prior choices on the between-trial heterogeneity in a logistic regression network meta-analysis.

Contributions to the Theory and Practice of Prior Elicitation
in Biopharmaceutical Research

by

Somer Blair, B.A., M.S.

A Dissertation

Approved by the 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 the Dissertation Committee

_____

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

_____

David J. Kahle, Ph.D., Co-Chairperson

_____

James D. Stamey, Ph.D.

_____

Jack D. Tubbs, Ph.D.

_____

Mark C. Kelly, Ph.D.

Accepted by the Graduate School
May 2017

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to the faculty and all those in the Department of Statistical Science at Baylor who create a family and a team. I could not ask for a better "statistics family" with which to share this accomplishment.

Thank you to Dr. John Seaman. I can always count on your constant encouragement and guidance. It seems that you always see excellence in me which has given me great confidence. You work selflessly for your students' best interest, and you are generous in your help. You are an exemplary mentor and someone that I will always hold in high regard.

Thank you to Dr. David Kahle. Your energy and passion are contagious. Thank you for sharing your brilliant ideas, your gift of teaching, and your caring advice. Thank you to Dr. Jack Tubbs and Dr. James Stamey, who are generously encouraging, supportive, and considerate.

Thank you to my fellow students for a challenging and lively atmosphere. Thank you, especially to my friend, Johanie, for your continuous comradery and encouragement. Thank you to my supportive and excellent friends and family outside of the department. What a blessing that there are too many to name!

Thank you to my in-laws, Jack and Gail, who have selflessly given their time and energy to see me reach my goal. Even though you did not raise me, you have shown as much pride and love for me as if you did. You never fail to support and encourage me with your wisdom and are always quick to offer generous help and guidance.

Thank you to my mom and late "Daddio", who have always seen the best in me. Who, when my older siblings informed me that I could not be a teacher, professional ice-skater, fire-fighter, and doctor all at the same time, responded with a gasp and, "Don't tell her that! Yes she can!" You have never failed to make evident your unconditional love

for me and pride in who I am. You instilled in me a positive attitude and confidence that I could accomplish whatever I set out to do, and you have helped me all along the way.

Thank you to my beloved, extraordinary children, James and George. You make me laugh and refresh my mind in the midst of the most difficult and frustrating problems. You remind me that I have been given abundant goodness and favor.

Thank you to my husband, Casey. It is hard for me to express how abundantly blessed and constantly proud I am to be your wife. You are my best friend, my rock, coach, and teammate. It seems that you see only the very best in me, though you often see me at my worst. You were the one that encouraged me to apply to graduate school, and you have never stopped encouraging me all the way to the end. You continuously reveal greatness, beauty, and strength in me that I sometimes cannot see. You believe that I am capable of anything, and I thank you that I have come to believe that, too. I am fervently excited to see the strengths you will bring out in our children.

Most of all, I am thankful to God, my greatest encourager and teacher, whose graciousness, mercy, and abundant generosity have been made evident in my life.

To my love and best friend, my husband, Casey

CHAPTER ONE

Introduction

Bayesian inference requires specification of a data model using a likelihood function, $\ell(\boldsymbol{\theta} \mid \mathbf{x})$, for some vector parameter, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and data vector $\mathbf{x} \in \mathcal{X}$. In addition, a joint prior distribution, $\pi(\boldsymbol{\theta})$, must be constructed. Then Bayes' theorem provides the posterior distribution on $\boldsymbol{\theta}$ given $\mathbf{x}$:

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}) \propto \ell(\boldsymbol{\theta} \mid \mathbf{x})\pi(\boldsymbol{\theta}). \tag{1.1}$$

Priors can be chosen to be more or less informative with a corresponding effect on the posterior. In this dissertation, we consider several aspects of prior construction, from elicitation of informative priors to unintended consequences of unnecessarily diffuse priors.

The literature on prior construction is vast. Typically, a prior is considered to be either vague or informative. Basic methods of construction are discussed in most texts on Bayesian inference. See, for example, Christensen et al. [10].

Relatively non-informative priors can be used. (See, for example, Robert [47] and Kass and Wasserman [32]). These can lead to difficulties with induced priors, as noted by Seaman et al. [29], especially in analyses with small sample sizes.

Elicitation of expert information and evaulation of resulting priors is critical. O'Hagan et al. [41] and Garthwaite et al. [23] provide nice overviews of this subject, including discussions on criteria for successful elicitations and the psychology behind eliciting opinions. Regarding the latter, Garthwaite et al. recommend eliciting information in terms of observables with which the experts are familiar. To this end, Bedrick et al. [4] introduce conditional means priors for regression models. Legedza & Ibrahim [34] and Chaloner et al. [7] discuss prior elicitation in clinical trials. The latter presents a graphical tool for doing so. Wu et al. [53] present sensitivity analyses for beta priors with inprecise expert

information in clincial trials. Combining expert opinion is discussed in [44], [45], and [41]. Historical data can also be utilized in prior construction. The power prior, introduced by Chen [9], incorporates historical data. Discussion can be found in [8] and [28].

1.1 Plan of Dissertation

The dissertation is organized as follows. In Chapter Two, we propose a prior elicitation method for use in parameteric proportional hazards models. When eliciting expert opinion, it is recommended to communicate in terms of observables with which the expert is most familiar. Therefore, we present methodology and simulated examples for obtaining a prior structure on the parameters of interest through information on median and mean survival times.

In Chapter Three, we consider power priors, which facilitate utilization of historical information in current experiments. A critical aspect of their use is the choice of the power parameter, a_0, which limits the influence of the historical data on the posterior. We propose methods for selection of a fixed a_0 through operational assessments based on implications of such choices. These methods utilize the prior effective sample size [36], the FDA guidelines on Bayesian analyses [2], and expert information.

We turn our attention to relatively non-informative priors in Chapter Four. Through simulated logistic datasets, we investigate the effect of diffuse prior choices in Bayesian network meta-analyses. We consider prior specifications regarding the between-trial heterogeneity and consequences of ignoring the covariate and/or the baseline in meta-regression. Our simulations suggest that caution should be taken when modeling small logistic datasets with few successes in meta-analyses.

In Chapter Five, we study the effect of expert misspecifications in prior elicitation with beta priors on the success probability in binomial data. Specifically, when the prior effective sample size [36] is used to assess priors in an analysis, we find that there can be

substantial sensitivity of priors to slight deviations in expert-elicited summaries. We propose a graphical guide for aide in the elicitation process to mitigate the sensitivity problem.

CHAPTER TWO

Prior Elicitation in Parametric Proportional Hazards Models

2.1 Introduction

Survival analysis consists of statistical techniques for modelling the time until the occurence of an event. The proportional hazards model (PHM) is commonly used in survival analysis where interest is in comparing the survival of two or more groups. The PHM assumes these groups have ratios of hazard rates which are not dependent on time, and hence have proportional hazards. We assume that the hazard rate follows a specific parametric form, which makes our model a parameteric PHM. We focus on prior elicitation in two models in particular, the exponential PHM and the Weibull PHM.

Expert opinion is necessary in many applications of the PHM, especially in exploratory and early phase research. In this regard, we consider informative priors on parametric PHM's with one dichotomous covariate, which indicates whether a patient is in the experimental or standard treatment group. We extend our methods to models with multiple categorical variables in Section 2.5. Following advice of many in the literature ([30], [35], [23]), we elicit information on observables with which experts are familar. We then convert this information to priors on the model's parameters of interest. We introduce notation and the basic PHM in Section 2.1.1. We provide the exponential and Weibull proportional hazards models in Sections 2.1.2 and 2.1.3, respectively. We continue in Section 2.2 with an illustration using prior specification through median survival times in the exponential PHM. Our methods are extended to the Weibull PHM in Section 2.3. We discuss parameterization of the Weibull model and implications in Section 2.3.1. We propose a prior elicitation method through median survival times in Section 2.3.2 and use simulated data to provide an example analysis. Section 2.4 includes construction of conditional means priors for the exponential PHM in Section 2.4.1 and the Weibull PHM in Section 2.4.2. In Section 2.5

we briefly discuss applicaton of our methods in models with multiple cateogrical variables and present an illustration in Section 2.5.1. We end the chapter with a discussion in Section 2.6.

2.1.1 Introduction of the Survival Model

We begin with a brief introduction to survival analysis in this section to establish notation and terminology. We also present our particular survival model.

Let T be a survival time with continuous density, $f(t)$. Here, T represents the time until an event, such as death or disease recurrence. The survival function, $S(t)$, is the probability of surviving up to or beyond time t:

$$S(t) = \text{Pr}(T \geq t).$$

In practice, we don't always observe t. Instead, for each patient i, we observe $\min(t_i, c_i)$ where c_i is a censored time. Censoring occurs when a death (or failure) time is not observed. We consider right-censoring in our model and account for it through the likelihood function. Details can be found in [33].

The hazard function $h(t)$ is the instantaneous rate of failure (or death) given survival to time t, defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{\text{Pr}\left\{t \leq T < t + \Delta t \mid T \geq t\right\}}{\Delta t} = \frac{f(t)}{S(t)}. \tag{2.1}$$

The proportional hazards model (PHM) has the form

$$h(t \mid \mathbf{z}) = h_0(t)\text{exp}\left(\boldsymbol{\beta}\mathbf{u}\right) \tag{2.2}$$

where $\mathbf{u}$ is a vector of covariates not dependent on t, and $h_0(t)$ is called the baseline hazard function.

There are several approaches to modelling $h_0(t)$, ranging from assuming a specific functional form for $h_0(t)$ to leaving $h_0(t)$ completely unspecified and using Dirichlet pro-

cess priors. We assume specific parametric baseline hazards, the exponential and Weibull, in our models.

Ibraham and Chen [27] discuss a piece-wise constant baseline hazard function function with gamma priors and with a power prior. Ibraham and Chen also use a gamma process prior on the baseline hazard and a parametric approach with a Weibull baseline hazard function in which non-informative priors are placed on the Weibull parameters. Christensen et al. [10] mention the use of an auto-regressive structure through a random walk prior for cubic B-splines on $h_0(t)$. They also model the baseline hazard function on sub-intervals through a piece-wise constant function.

We assume that $h_0(t)$ follows a particular parametric form, which makes our model a parametric PHM. The coefficient-vector, $\boldsymbol{\beta}$, usually does not include an intercept term. Inclusion of an intercept would produce a constant exponential term that could be absorbed into $h_0(t)$. We focus on PHM's with one binary covariate, z:

$$h(t \mid z) = h_0(t)\text{exp}\left(\beta z\right). \tag{2.3}$$

We consider Weibull and exponential PHM's, where $z = 0$ denotes a standard treatment group and $z = 1$ the experimental group. Often, clinical interest is in the hazard ratio. Let h_1 and h_2 be the hazard functions for the experimental and standard treatment groups, respectively. Then, the hazard ratio is

$$HR = \frac{h_1(t)}{h_2(t)} = \frac{h_0(t)\text{exp}(\beta \times 1)}{h_0(t)\text{exp}(\beta \times 0)} = \text{exp}(\beta). \tag{2.4}$$

Our methods can be extended to include multiple categorical covariates, as well. We present this in Section 2.5.

2.1.2 Exponential Proportional Hazards Model

A commonly-used parametric form for the baseline hazard is a constant function corresponding to the exponential PHM. Recall the general form of the PHM in (2.2). We

specify an exponential baseline hazard function

$$h_0(t) = \gamma. \tag{2.5}$$

For a single dichotomous covariate, z, this leads to the exponential PHM

$$h(t \mid z) = \gamma \text{exp}(\beta z). \tag{2.6}$$

with survival function

$$S(t) = \text{exp}\left\{ -\int_0^t \gamma \text{exp}(\beta z) du \right\} = \text{exp}\left\{ -t\gamma \text{exp}(\beta z) \right\} \tag{2.7}$$

and corresponding probability density function

$$f(t) = \gamma \text{exp}(\beta z) \text{exp}\left\{ -t\gamma \text{exp}(\beta z) \right\}. \tag{2.8}$$

This is an exponential density with rate $\gamma \exp(\beta z)$. Let $\boldsymbol{\delta}$ be a vector of censoring indicators where δ_i equals one if the ith death time is observed and zero otherwise. Then the likelihood corresponding to the exponential PHM in (2.6) is:

$$\begin{aligned} \ell(\beta, \gamma \mid \mathbf{t}, \mathbf{z}, \boldsymbol{\delta}) = \prod_{i=1}^{n} &\left\{ \gamma \text{exp}(\beta z_i) \text{exp}\left[-t_i \gamma \text{exp}(\beta z_i) \right] \right\}^{\delta_i} \\ &\times \left\{ \text{exp}\left[-t_i \gamma \text{exp}(\beta z_i) \right] \right\}^{1-\delta_i} \end{aligned} \tag{2.9}$$

where $\mathbf{t}$ is a vector of n survival times and $\mathbf{z}$ is the corresponding vector of dicotomous covariate values.

2.1.3 Weibull Proportional Hazards Model

The Weibull PHM is a generalized form of the exponential PHM. Specifying a Weibull baseline hazard function in (2.2),

$$h_0(t) = \lambda r t^{r-1}, \tag{2.10}$$

yields, for dichotomous z, the Weibull PHM:

$$h(t \mid z) = \lambda r t^{r-1} \text{exp}(\beta z). \tag{2.11}$$

The survival function at time t is

$$S(t) = \exp\left\{-t^r \lambda \exp(\beta z)\right\}. \tag{2.12}$$

The probability density function of t is

$$f(t) = r t^{r-1} \lambda \exp(\beta z) \exp\left\{-t^r \lambda \exp(\beta z)\right\}. \tag{2.13}$$

This is a Weibull density with scale determined by $\lambda \exp(\beta z)$ and shape parameter r. Let $\delta = 1$ if the death time is observed and zero otherwise. The likelihood for this model is

$$\begin{aligned} \ell_w(\beta, r, \lambda \mid \mathbf{t}, \mathbf{z}, \boldsymbol{\delta}) = \prod_{i=1}^{n} &\left\{ r t_i^{r-1} \lambda \exp(\beta z_i) \exp\left[-t_i^r \lambda \exp(\beta z_i)\right] \right\}^{\delta_i} \\ &\times \left\{ \exp\left[-t^r \lambda \exp(\beta z)\right] \right\}^{1-\delta_i}, \end{aligned} \tag{2.14}$$

where **t**, **z**, and $\boldsymbol{\delta}$ are defined as before.

The exponential PHM is the Weibull PHM with the shape parameter, r in (2.11), equal to one.

2.2 Priors for the Exponential PHM through the Median Survival Time

For a Bayesian analysis, the joint posterior distribution for parameters of interest is proportional to the product of the likelihood ((2.9) or (2.14)) and a joint prior. In practice, this computation will require Markov chain Monte Carlo (MCMC) methods and will be implemented in a software package such as OpenBUGS or STAN.

In this section, we present methodology for eliciting priors on exponential PHM parameters through expert information on the median survival time. An illustration of the prior construction and an example analysis are provided in Section 2.2.1.

Given expert information on the median survival time, we wish to obtain a prior structure on exponential PHM parameters γ and β. Let t_{m_0} and t_{m_1} be the median survival times for the standard and experimental treatment groups, respectively. By definition, these median survival times satisfy

$$S(t_{m_0} \mid z = 0) = 0.5 \tag{2.15}$$

and

$$S(t_{m_1} \mid z = 1) = 0.5. \tag{2.16}$$

We elicit expert opinion on observables t_{m_0} and t_{m_1} and represent this opinion with probability distributions. This induces distributions on the parameters of interest, γ and β. To construct the priors on the median survival times, we obtain modes and percentiles from the expert.

Suppose the expert provides most likely values and high values for the median survival times above. Let $t^*_{m_1}$ and $t^*_{m_0}$ be the most likely values of t_{m_1} and t_{m_0}, respectively. These values will be treated as the modes of their respective probability distributions. Let the high values of the median survival times for the standard and experimental treatment groups be denoted by u_{m_0} and u_{m_1}, respectively. Note that low values could be used instead. The expert should think of the high value, u_{m_i}, as an unusually large value of the median survival time, such that there is small probability of the median survival being greater than u_{m_i}. We treat high values as upper percentiles of gamma distributions. We recommend not eliciting extreme percentiles, and the 90^{th} should be the highest considered [41].

We begin by specifying probability distributions on t_{m_0} and t_{m_1}. We choose gamma priors

$$t_{m_0} \sim \text{Gamma}(k_0, \theta_0) \tag{2.17}$$

and

$$t_{m_1} \sim \text{Gamma}(k_1, \theta_1). \tag{2.18}$$

In our development, these denote the probability density functions given by

$$f(t_{m_i}) = \frac{1}{\Gamma(k_i)\theta^{k_i}} t_{m_i}^{k_i - 1} \exp\left(-\frac{t_{m_i}}{\theta_i}\right) \quad i = 0, 1 \tag{2.19}$$

where $\Gamma(\cdot)$ is the incomplete gamma function. Denote the corresponding cummulative density function by $F(t)$. We construct the priors to have modes at the elicited values $t^*_{m_0}$

and $t^*_{m_1}$. For the parameterization in (2.19) we can write

$$t^*_{m_i} = (k_i - 1)\theta_i, \quad i = 0, 1. \tag{2.20}$$

Using the expert's high values, u_{m_0} and u_{m_1}, we set

$$F(u_{m_i}) = \frac{1}{\Gamma(k_i)}\gamma\left(k_i, \frac{u_{m_i}}{\theta_i}\right) = \xi, \quad i = 0, 1. \tag{2.21}$$

where the choice of ξ determines the informativeness of the gamma prior. We will see that our proposed prior structure often results in highly informative induced priors. We ultimately recommend $\xi = 0.75$ for our proposed methods. This results in a system of equations defined by (2.20) and (2.21) with which we can solve for the gamma parameters k_i and θ_i, $i = 0, 1$ in (2.17) and (2.18). This can be done numerically. In many applications, more is known about the standard treatment than the experimental treatment. In that case, t_{m_0} should have a smaller variance.

Together, the definitions of the median survival times in (2.16) and (2.15) and the survival function in (2.7) yield the following relationships:

$$t_{m_0} = \frac{\ln(2)}{\gamma} \tag{2.22}$$

and

$$t_{m_1} = \frac{\ln(2)}{\gamma \exp(\beta)}. \tag{2.23}$$

Therefore, the parameters of interest have direct relationships with the median survival times through

$$\gamma = \frac{\ln(2)}{t_{m_0}} \tag{2.24}$$

and

$$\exp(\beta) = \frac{\ln(2)}{\gamma t_{m_1}}. \tag{2.25}$$

Equation (2.24) induces a prior on γ through t_{m_1}, and a prior is induced on β through (2.25).

2.2.1 *Illustration of Prior Construction*

The purpose of our development in this chapter is to provide a means to use expert opinion in fitting PHM's. To the degree we accurately represent the opinon of a well-informed expert, we expect posterior estimates to perform well compared to those obtained using methods uninformed by prior information, such as maximum likelihood. However, for completeness, we now provide an illustration of the use of such priors.

As an example, suppose the expert believes that the median survival time for the experimental treatment is most likely 75 weeks and is less than 105 weeks. The median survival time for the standard group is not quite as high as the experimental group, but the expert is more certain about the survival times of the former. Suppose further that the expert believes the median survival time for the standard treatment is most likely 67 weeks and is less than 90 weeks. Thus, the elicited quantities for the standard treatment are $t^*_{m_0} = 67$ and $u_{m_0} = 90$. For the treatment group, these are $t^*_{m_1} = 75$ and $u_{m_1} = 105$.

The elicited values of $t^*_{m_i}$ and u_{m_i} are treated as modes and 75^{th} percentiles of the gamma distributions in (2.17) and (2.18). O'Hagan et al. [41] discuss reasons against using extreme percentiles, such as the 90th or 95th. Thus, we have the following gamma priors on t_{m_0} and t_{m_1}:

$$t_{m_0} \sim \text{Gamma}(k_0 = 9.2, \theta_0 = 8.1) \tag{2.26}$$

and

$$t_{m_1} \sim \text{Gamma}(k_1 = 7.6, \theta_1 = 11.4). \tag{2.27}$$

We can examine the priors induced on β and γ by these choices via simulation, using (2.24) and (2.25). These induced priors are exhibited in Figure 2.1. Note that these priors can be approximated using independent distributions, which we later discuss.

If possible, one should include induced priors on other parameters of clinical interest, such as the hazard ratio, HR. This facilitates feedback from the expert, allowing for an iterative process of fitting and checking. In this case, the mode (and mean) of the prior on

HR is 1. Therefore, the expert's opinion has implied that the experimental treatment may not be much different than the standard treatment.

For a posterior analysis, we specify the gamma priors on the median times and the relationships in (2.24) and (2.25) to induce the priors in OpenBUGS. The priors representing expert opinion on the median survival times and the induced priors on the parameters of interest are can be viewed via simulation and are displayed in Figure 2.1. Note that simulation of the induced priors is not required for implementation in packages such as OpenBUGS. Though, observing the priors is helpful in collaboration with the expert and for sensitivity analyses.

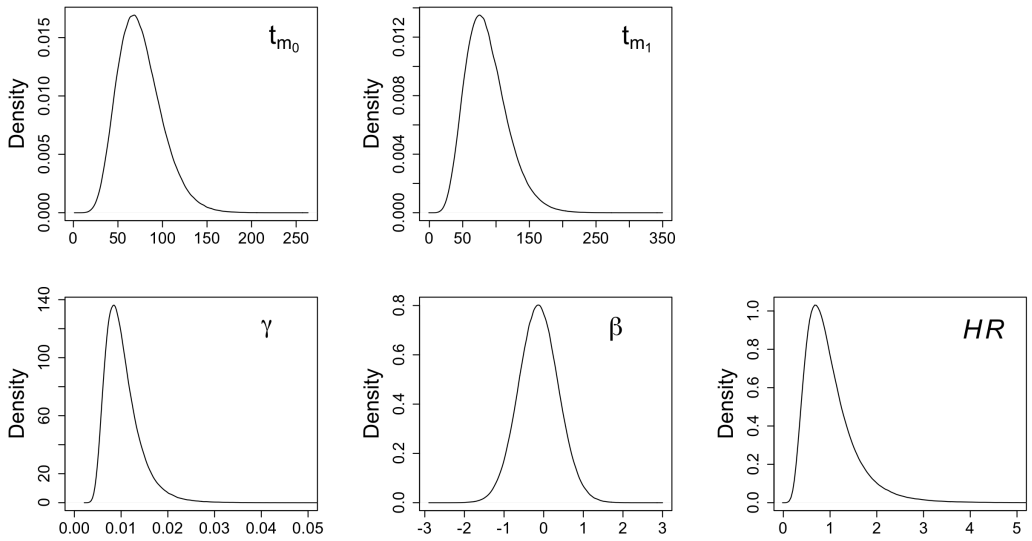


Figure 2.1: Gamma densities on t_{m_0} and t_{m_1} with induced prior densities on γ, β and HR.

To complete the example, suppose we have 50 exponential survival times corresponding to the PHM in (2.6), including 3 censored times, with parameters $\gamma = 0.0107$ and $\beta = -0.268$ corresponding to true median survival times of $t_{m_0} = 65$ and $t_{m_1} = 85$. The survival times are plotted in Figure 2.2. The value of the hazard ratio $HR = \exp(\beta) = 0.765$. We have assumed our expert's knowledge is accurate and precise. That is, the true median times are not far from the corresponding expert-elicited most likely values. In fact, the true values are near the center of their respective gamma priors.

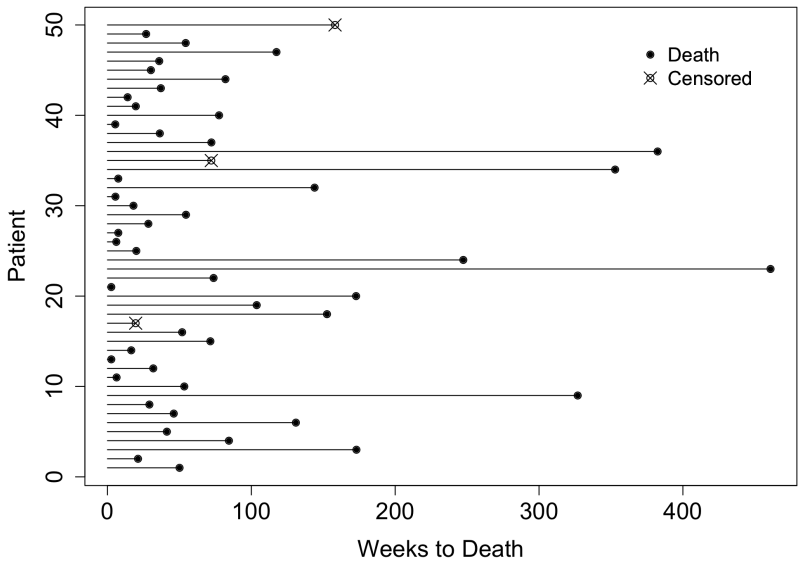


Figure 2.2: Simulated exponential survival times.

We fit this model using OpenBUGs by running 50,000 iterations following a 10,000 iteration burn-in. No convergence issues were encountered. The means, 2.5^{th}, and 97.5^{th} percentiles of the posterior distributions are shown in Table 2.1 along with the frequentist estimates. The frequentist estimates are found using the `R` package, `SurvRegCensCov`. This package computes maximum likelihood estimates (MLE) numerically by maximizing the Cox partial likelihood [33]. The 95% confidence interval is computed for HR using the asymptotic normality of the MLE (for details see Chapter eight of [33]), and the standard errors (SE) are computed using the delta method (see Chapter twelve of [33]).

Table 2.1: Posterior estimates and maximum likelihood estimates of the exponential PHM parameters.

	Truth	Posterior Estimates			Frequentist Estimates	
		Mean	2.5%	97.5%	MLE	SE
γ	0.0107	0.012	0.007	0.017	0.012	0.0026
β	-0.268	-0.190	-0.702	0.318	-0.216	0.292
HR	0.765	0.859	0.499	1.383	0.806	0.235

The Bayesian estimates contain the true values within their respective 95% posterior credible intervals. Also, each of the true values are within one standard error of their respective MLE's. The posterior mode is slightly farther from the truth than the MLE's for β and HR, but the two are comparable. The posterior mean estimates γ better than the MLE, in this example. Figure 2.3 displays the prior and posterior densities of the parameters with solid dots indicating respective MLE's.

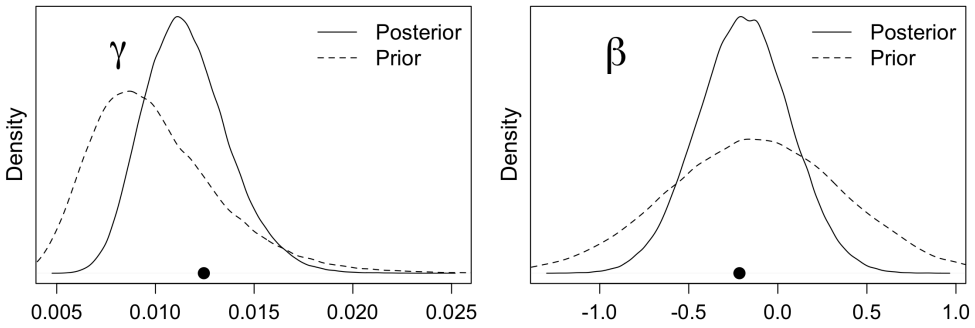


Figure 2.3: Posterior densities of β and γ with solid dots at corresponding MLE's.

In Table 2.1, the posterior mean and MLE are identical for γ. For β and HR, the posterior mean and MLE are comparable. In this example, the MLE is slightly closer to the true value.

It is common in the literature to treat the parameters of the PHM as independent in prior construction — see the discussion in Section 2.7. Here, we take γ and β to be independent with priors

$$\gamma \sim \text{LN}(-4.65, 0.32^2) \tag{2.28}$$

and

$$\beta \sim \text{N}(-0.127, 0.502^2),$$

where $x \sim \text{LN}(\mu, \sigma^2)$ means that $e^x \sim \text{N}(\mu, \sigma^2)$.

Specifying the approximate independent priors in OpenBUGS, we obtain posterior means of 0.012 and –0.14 for γ and β, respectively. The corresponding 95% credible intervals are: $\lambda \in (0.008, 0.016)$ and $\beta \in (-0.62, 0.33)$. The posterior results in Table 2.1,

using dependent priors, have means closer to the truth and wider 95% credible intervals. However, they are not far from the posterior results found using independent priors.

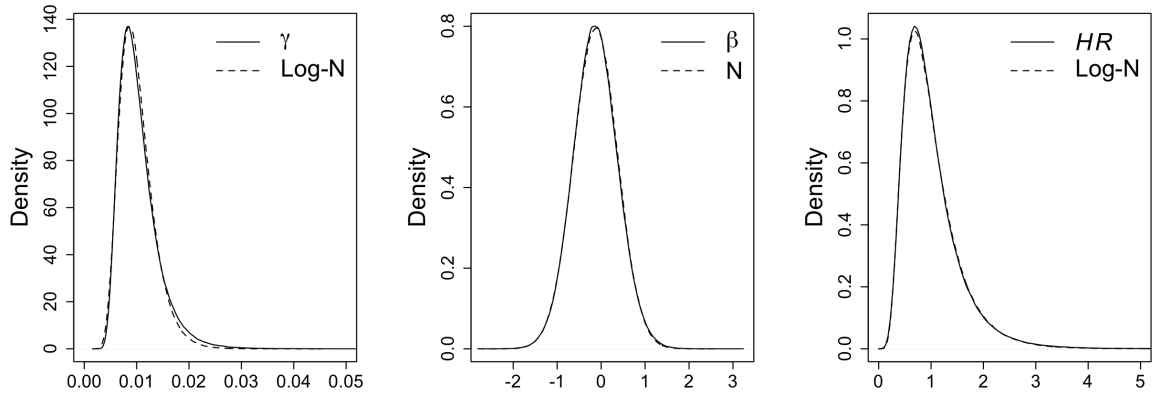


Figure 2.4: Induced densities on γ, β and HR with approximate independent priors (dotted lines).

We recommend specifying the functional relationships between the median times and the PHM parameters directly in OpenBUGS. The additional computation time is negligable, and this method automatically accounts for dependence among the parameters through the relationships in (2.24) and (2.25). Approximating the induced priors with independent distributions does not account for this dependency. Also, we have encountered induced priors that are difficult to fit with independent distributions.

2.3 Priors for the Weibull PHM

In this section, we consider priors on parameters for the Weibull PHM. We detail properties and parameterizations of the model, present prior formulation methods through elicitation on observables, and provide an example analysis.

2.3.1 Parameterization and Properties of the Weibull PHM

Exponential survival times are, of course, a special case of the Weibull distribution. General properties of this distribution can be found, for example, in Section 2.2 of [46].

As we shall see, the choice of parameterization for the Weibull is critical to effective prior elicitation. Recall from Section 2.1.3 that the survival times corresponding to our Weibull PHM have density (2.13). In general, the Weibull is a three-parameter distribution including a shift parameter, η. The three-parameter Weibull probability density determined by shift, scale, and shape parameters, η, ξ, and r, respectively, is

$$f(t \mid \eta, \xi, r) = r\xi(t-\xi)^{r-1} \exp[-\xi(t-\xi)^r], \quad t \geq \xi. \tag{2.29}$$

Increasing or decreasing η causes the density curve to shift right or left, respectively. Because of the shifting effect of η, there is no lifetime t which is smaller than η. For this reason, η is also called a minimum lifetime [46]. We take the shift parameter, η, to be zero and denote this distribution, using scale and shape parameters, by $t \sim \text{Weibull}(\xi, r)$. By taking $\eta = 0$, we assume that the survival times begin at zero, instead of later in a time period.

Expert information can be used in the prior elicitation process to construct a distribution on r as follows. When the shape parameter, r, equals one the Weibull is equivalent to the exponential distribution. When $r < 1$ or $r > 1$, the hazard rate is monotonically decreasing or increasing, respectively. The shape parameter is also called the Weibull-slope because r can be thought of as the slope of the CDF [46]. For example, suppose $r = 5$. Then, as time increases by 1 unit, the CDF increases by 5%. In other words, if $r = 5$, as time increases by 1 unit, survival decreases by 5%. Thus, supposing that t is in months, we can ask the expert to provide the expected percent survival decrease after one month.

Increasing and decreasing the scale parameter, ξ, will cause the variation of t to become larger and smaller, respectively. In our development, we focus on the Weibull PHM which yields Weibull survival times t with scale ξ, equal to $\lambda \exp(\beta z)$ and shape parameter r.

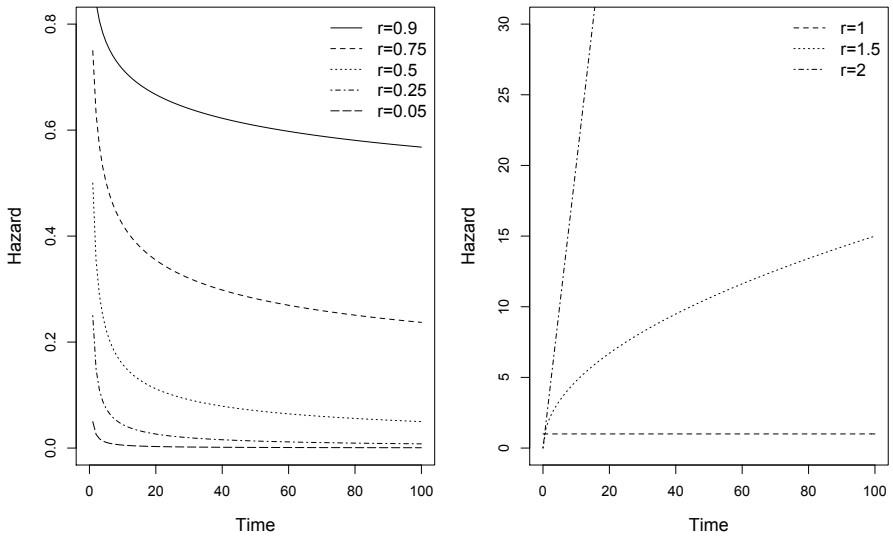


Figure 2.5: Weibull density with differing values of the shape parameter

2.3.2 Prior Elicitation Through Median Survival Times

Similar to Section 2.2, we use median survival times to inform choices of probability distributions on parameters of interest. We wish to induce priors on β, r, and λ for the Weibull PHM in (2.11).

The definitions of the median survival times t_{m_0} and t_{m_1} in (2.16) and (2.15) yield the following relationships for the Weibull parameters:

$$r = \frac{\log[\log(2)/\lambda]}{\log(t_{m_0})} \tag{2.30}$$

and

$$\exp(\beta) = \frac{\log(2)}{\lambda t_{m_1}^r}. \tag{2.31}$$

The parameter λ acts as a scaling parameter and has no operational interpretation. However, the choice of λ does have repercussions in the model, especially with respect to the shape parameter, r. In particular, if $\lambda \exp(\beta) > \log(2)$ and $t_{m_0} \geq 1$, then r will be negative. Indeed, t_{m_0} is typically larger than 1. However, r must be positive, and so must

the scale, denoted ξ in (2.29), which requires the bounds

$$0 < \lambda \exp(\beta) < \log(2).$$

Let d be an unusually large value of the hazard ratio in practical settings. Then, we bound λ by [1]

$$0 < \lambda < \frac{\log(2)}{d}. \tag{2.32}$$

The choice of d should be determined via a sensitivity analysis. If d is too large, the posterior density on λ will "pile up" at the upper bound. This occurs when the data suggests small values of HR. However, a value of d that is too small results in unreasonably precise priors on β and HR. Other than the required bound, there is no operational interpretation of λ. We give λ a uniform distribution between 0 and $\log(2)/d$:

$$\lambda \sim \mathrm{U}\left(0, \frac{\log(2)}{d}\right). \tag{2.33}$$

More generally, a beta distribution on the interval in (2.32) could be selected. To this end, we use a four-parameter beta distribution. For a random variable x bounded by p and q with density function

$$f(x) = \frac{(x-p)^{a-1}(q-x)^{b-1}}{B(a,b)(q-p)^{a+b-1}}, \tag{2.34}$$

we write $x \sim \text{Beta}_{[p,q]}(a,b)$. Thus, using the interval in (2.32) as support, we take

$$\lambda \sim \text{Beta}_{[0,\log(2)/d]}(a,b). \tag{2.35}$$

To illustrate the uniform and more general beta priors, we let $d = 3$ so that λ is bounded between 0 and approximately 0.23. We choose values of shape parameters a and b that concentrate the beta prior on small values of λ. This parameterization yields potential priors for λ of $\mathrm{U}(0, \log(2)/3)$ or the beta prior

$$\lambda \sim \text{Beta}_{[0,\log(2)/3]}(1.0, 1.2). \tag{2.36}$$

[1] Note that if the median survival time is less than 1, then λ must be larger than $\log(2)/\exp(\beta)$.

Figure 2.6 shows both priors for graphical comparison. The beta prior is slightly more informative than the uniform, giving more probability to smaller values of λ. Increasing b will increase the peak at the smaller λ values which is of interest if we desire smaller prior values of λ.

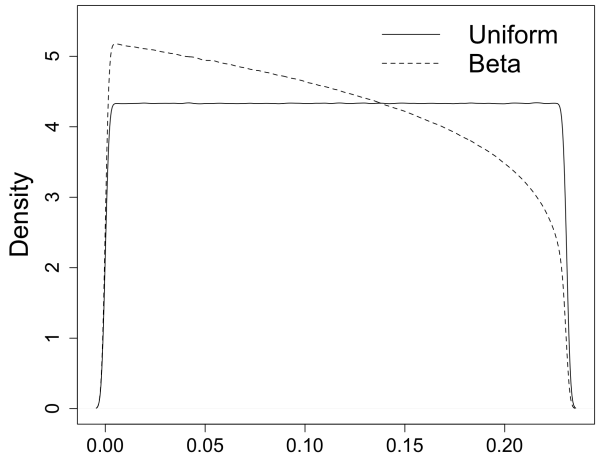


Figure 2.6: Comparison of a uniform versus beta distribution as a potential prior on λ.

Figure 2.7 displays an example of induced priors on the Weibull parameters using the uniform (2.33) and beta (2.36) priors on λ. Interestingly, the beta prior induces a prior on r that is slightly more diffuse than that induced by the uniform distribution. The difference is minimal, and the choice between these priors has little effect on the induced priors on β and HR. We choose a uniform prior on λ for our example below.

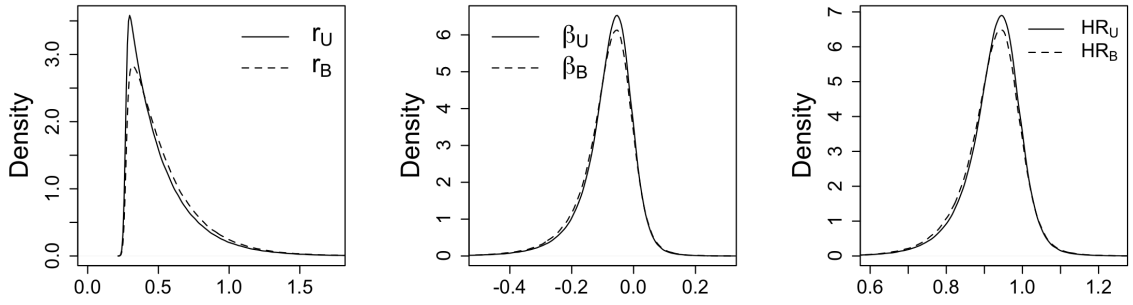


Figure 2.7: Comparison of induced priors using the uniform (subscript "U") versus the beta (subscript "B") prior on λ.

As in Section 2.2, through expert elicited modes and upper percentiles, we give t_{m_0} and t_{m_1} gamma priors to represent an expert's opinion. We use a uniform prior on the nuisance parameter λ and find induced priors on the Weibull PHM parameters. The median survival time for the standard group, t_{m_0}, and the parameter λ induce a prior on r through (2.30). The induced prior on r and the priors on t_{m_1} and λ induce a prior on HR through (2.31).

For an example, we simulate 50 Weibull survival times, including 8 censored times, with parameter values $r = 0.45$, $\beta = -0.13$, nuisance parameter $\lambda = 0.08$, and $HR = \exp(\beta) = 0.88$. The relationships in (2.30) and (2.31) imply that the corresponding true median survival times are $t_{m_0} = 120$ and $t_{m_1} = 160$, in weeks. The simulated data is plotted in Figure 2.8 along with corresponding theoretical hazard functions for each group.

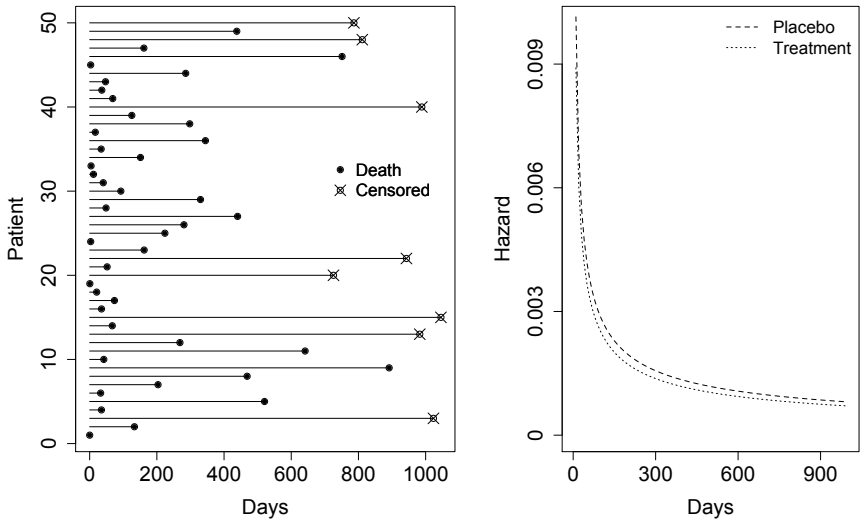


Figure 2.8: Simulated survival times and corresponding theoretical hazard functions.

Suppose an expert provides most likely values of t_{m_0} and t_{m_1} of 100 and 125 weeks, respectively. Also, the expert believes that t_{m_0} and t_{m_1} are no greater than 120 and 165, respectively. Using the notation from Section 2.2.1, we have $t^*_{m_0} = 100$ and $u_{m_0} = 120$ for the standard treatment group, and $t^*_{m_1} = 125$ and $u_{m_1} = 165$ for the experimental group.

For this example, we take $d = 5$ to be an usually large value of the hazard ratio, which yields an upper bound of approximately 0.14 on λ. Using $d = 6$, which bounds λ below 0.12, would cause the posterior density on λ to "pile up" at the upper bound of 0.12. Using $d = 4$, which bounds λ below 0.17, induces in an overly informative prior on β. Thus, we have the following prior distributions:

$$\begin{aligned} t_{m_0} &\sim \text{Gamma}(k_0 = 14.37, \theta_0 = 7.47), \\ t_{m_1} &\sim \text{Gamma}(k_1 = 8.23, \theta_1 = 16.59), \end{aligned} \tag{2.37}$$

and

$$\lambda \sim \text{U}\left(0, \frac{\log(2)}{5}\right).$$

To view the induced prior structure on the parameters of interest, we simulate 100,000 values of t_{m_0}, t_{m_1}, and λ from their respective prior distributions in `R`. Using the functional relationships in (2.30) and (2.31) we compute 100,000 values of r, β, and HR. The resulting induced kernel density estimates are shown in Figure 2.9.

The priors on the median survival times induced fairly informative priors on β, r and HR. This is expected, but in practice, this might be an important source of feedback for the expert. If less informative priors are desired, the high values, u_{m_0} and u_{m_1}, can be set to a lower percentile. Setting the expert's most likely value to the median of the gamma distribution, rather than the mode, results in a less informative prior on β for this example. At times, the value of d can be increased to inflate the prior on β, as well.

For a posterior analysis, we ran one million iterations in OpenBUGS after a burn-in of 40,000. There was no evidence of convergence problems. The estimates are shown in Table 2.2. The means, 2.5^{th} and 97.5^{th} percentiles of the posterior distributions are shown along with the frequentist estimates. The MLE's are found numerically in `R` by maximizing the Cox partial likelihood. The 95% confidence interval is shown for HR. This was computed with the `R` package `SurvRegCensCov` using the asymptotic normality of the MLE (for details see Chapter eight of [33]). The package `SurvRegCensCov` also computes the standard errors using the delta method (see Chapter twelve of [33]).

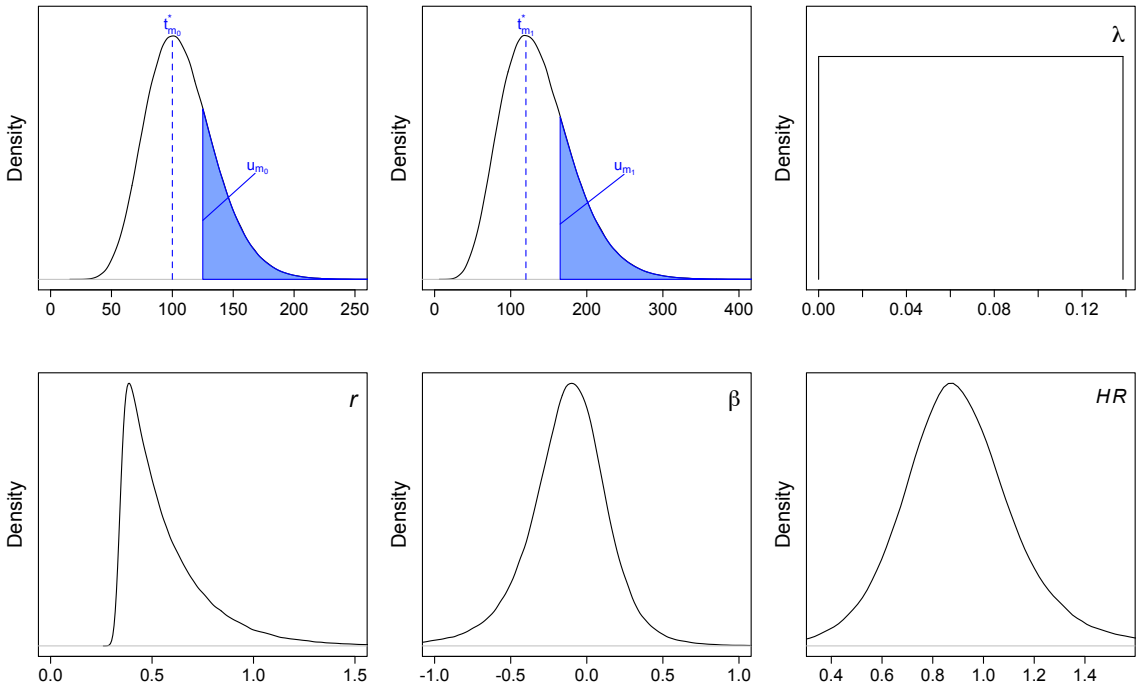


Figure 2.9: Figure: (Top: from left to right) Gamma densities on t_{m_0} and t_{m_1} and Uniform density on λ with (Bottom: from left to right) induced prior densities on r, β and the hazard ratio $\exp(\beta)$.

Table 2.2: Posterior results and frequentist estimates for the Weibull PHM parameters.

	Truth	Posterior Estimates			Frequentist Estimates	
		Mean	2.5%	97.5%	MLE	SE
λ	0.08	0.059	0.027	0.105	0.040	0.019
r	0.45	0.538	0.417	0.672	0.584	0.076
β	-0.13	-0.178	-0.534	0.172	-0.222	0.310
HR	0.88	0.851	0.586	1.187	0.800	C.I. $= (0.46, 1.47)$

The true values are contained in their respective 95% posterior credible intervals. Also, the MLE's all estimate the truth within two standard errors. Given the use of our informative priors, it is not surprising that the posterior means for β and HR are closer to the truth than the MLE's. The posterior densities of the parameters are displayed in Figure 2.10.

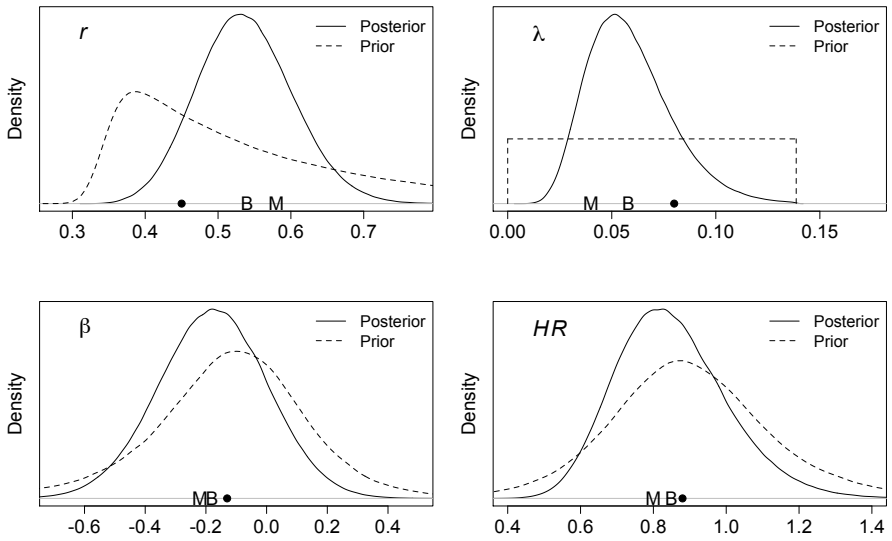


Figure 2.10: Posterior densities on HR, β , λ, and r. Solid dots, "M"'s, and "B"'s indicate true values, MLE's and posterior means, respectively.

As we discussed in Section 2.2.1, the PHM parameters are often treated as independent. We note those in the literature that do so in Section 2.6. To this end, we can approximate the induced priors in Figure 2.9 with independent priors on r and β. Here, we give r and β a shifted exponential and a normal prior, respectively. Specifically, we take

$$\beta \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2), \tag{2.38}$$

and, for $b > 0$, we use $r - b \sim \mathrm{Exp}(\xi)$. In effect, we shift the exponential to have support (b, ∞). We denote the shifted exponential by $\mathrm{Exp}(\xi, b)$. For our approximations, we use $\mu_\beta = -0.1$, $\sigma_\beta = 0.23$, $b = 0.35$, and $\xi = 4.33$. The resulting approximate densities are shown in Figure 2.11, revealing good fits for all, although the approximate normal on β truncates the tails of the true induced prior. Other approximating distributions were also considered. In particular, among the distributions that allow for negative support, the t-distribution was not sufficiently concentrated, and the skew-normal truncated one of the tails depending on the direction of skewness.

In our example, the normal approximation on β is more concentrated than the actual induced density. This is especially of concern because our priors are already highly infor-

mative. Therefore, appropriate caution should be taken with the induced prior structure. A prior to posterior sensitivity analysis is recommended. We repeat the example analysis using the approximate independent priors in (2.38). This results in posterior means of 0.06, 0.54, and –0.15 for λ, r, and β, respectively. The corresponding 95% credible intervals are: $\lambda \in (0.02, 0.11)$, $r \in (0.41, 0.68)$, and $\beta \in (-0.51, 0.21)$. Recall that the results in Table 2.2 were found by inducing the priors implicitly in OpenBUGS. Our results using the independent prior structure slightly deviate from those in Table 2.2. The deviation here is greater than the deviation between the two methods in the exponential PHM case in Section 2.2.

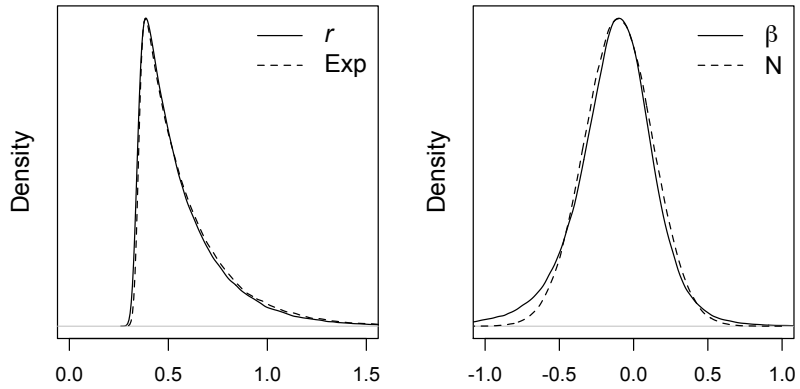


Figure 2.11: Induced priors on r and β with approximate independent densities (dotted lines).

2.4 Conditional Means Priors

A natural progression from eliciting prior information on the median survival time is a conditional means priors development for our proportional hazards model. Conditional means priors (CMP)rely on expert opinion about observables. It is prefereable to elicit information on observables which are on the same scale as the data (Tsutakawa and Lin (1986), [23], [4]). Given p predictor variables, we condition the mean on p locations in the predictor space. This yields p conditional means on which we elicit expert information. For more detail see [4].

The Weibull and exponential are, of course, skewed distributions. O'Hagan et al. note in Chapter five of [41] that expert's may have trouble specifying means, rather than medians, in skewed distributions. Therefore, caution must be taken when using conditional means priors in this context. For one, the expert should fully understand the difference between the mean and median and be comfortable specifying means for skewed data.

As before, suppose we have one binary predictor variable z. Let m_0 and m_1 be the conditional mean survival times for the standard treatment group and the experimental treatment group, respectively. The conditional means are defined as

$$m_0 = \mathrm{E}[t \mid z = 0] \tag{2.39}$$

and

$$m_1 = \mathrm{E}[t \mid z = 1]. \tag{2.40}$$

We use probability distributions on m_0 and m_1 to model information elicited from an expert. Distributions on m_0 and m_1 induce priors on parameters of interest from the PHM. In Section 2.4.1, we illustrate prior construction for the exponential PHM. We present an example analysis for the Weibull PHM in Section 2.4.2.

2.4.1 CMP for Exponential PHM

As in Section 2.1.2, suppose that we have an exponential PHM:

$$h(t \mid z) = \gamma \mathrm{exp}(\beta z). \tag{2.41}$$

The mean of the exponential distribution is equal to the inverse of its rate. Therefore, the conditional means are

$$m_0 = \mathrm{E}[t \mid z = 0] = \gamma^{-1} \tag{2.42}$$

and

$$m_1 = \mathrm{E}[t \mid z = 1] = \{\gamma \exp(\beta)\}^{-1}. \tag{2.43}$$

We elicit expert information on m_0 and m_1 in a way similar to that we used with median survival times. Specifically, we elicit the most likely mean value, m_i^*, as well as a high value, m_i^u, $i = 0, 1$. Given the expert's most likely values and high values for the means, we specify gamma distributions as priors for m_0 and m_1. The gamma distributions induce priors on γ and β through the relationships

$$\gamma = m_0^{-1} \tag{2.44}$$

and

$$\beta = \log\left(\frac{1}{\gamma m_1}\right). \tag{2.45}$$

To illustrate the prior elicitation, assume an expert believes that the most likely mean survival times for the standard and experimental treatment groups, respectively, are $m_0^* = 67$ and $m_1^* = 75$ weeks. Further, the expert is fairly certain that the the mean survival times are no greater than $m_0^u = 90$ and $m_1^u = 105$ weeks for the standard and experimental treatment groups, respectively.

Similar to the approach using medians in Sections 2.2 and 2.3.1, we treat the most likely values and high values as modes and 75^{th} percentiles, respectively, of gamma distributions. The expert opinion is represented by gamma distributions on m_0 and m_1:

$$m_0 \sim \text{Gamma}(k_0 = 9.2, \theta_0 = 8.1) \tag{2.46}$$

and

$$m_1 \sim \text{Gamma}(k_1 = 7.6, \theta_1 = 11.4). \tag{2.47}$$

The priors in (2.46) induce priors on γ and β through (2.44). We find the induced priors emperically in `R`. Simulating one million m_0 and m_1 values from their gamma distributions and specifying in `R` the functional relationships in (2.44) gives the kernel density plots in Figure 2.12. Visualizing the induced priors may be of use during elicitation. Comparison of priors with resulting posteriors is also a critical aspect of model checking. To this end, we

can approximate the prior densities in Figure 2.12 with independent distributions, displayed as dotted lines in Figure 2.12. In this case, the approximate priors are

$$\begin{aligned} \gamma &\sim \text{LN}(-4.28, 0.325^2); \\ \beta &\sim \text{N}(-0.127, 0.5^2). \end{aligned} \tag{2.48}$$

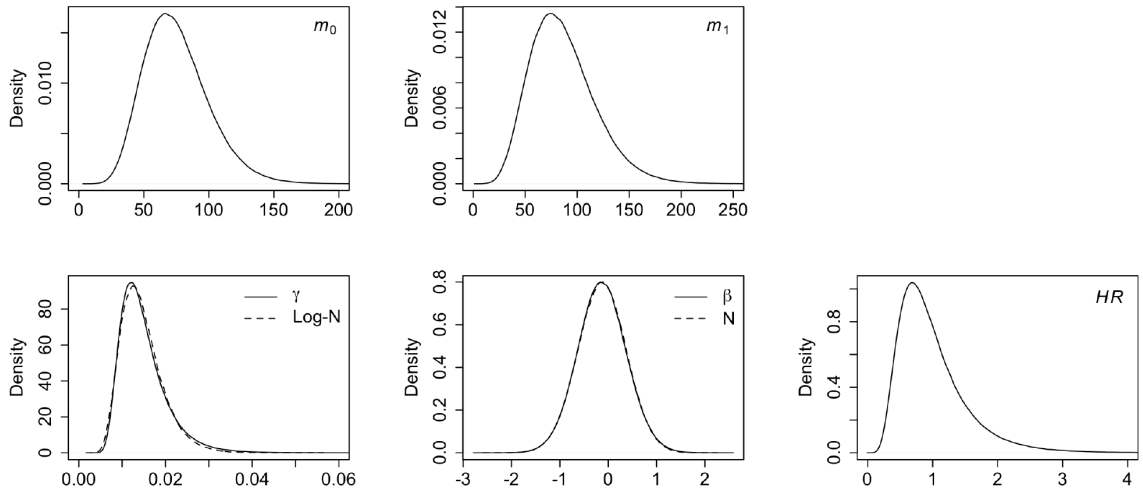


Figure 2.12: Priors representing expert opinion (top row) and resulting induced priors (bottom row).

To perform a Bayesian analysis via OpenBUGS, we recommend specifying the gamma priors on m_0 and m_1 in (2.46) along with their functional relationships to the parameters of interest in (2.44) in the model statement.

2.4.2 CMP for Weibull PHM

In this section, we apply conditional means priors to the Weibull PHM. As in Section 2.1.3, we have

$$h(t \mid z) = \lambda r t^{r-1} \exp(\beta z). \tag{2.49}$$

Therefore, the conditional mean survival times are defined as

$$m_0 = \text{E}[t \mid z = 0] = \lambda^{-1/r} \Gamma\left(1 + 1/r\right) \tag{2.50}$$

and

$$m_1 = \mathrm{E}[t \mid z = 1] = \{\lambda \exp(\beta)\}^{-1/r} \Gamma\left(1 + 1/r\right). \tag{2.51}$$

We cannot solve for r explicitly in (2.50). However, we can use the `R` function `rootsolve` to solve for r using the Newton-Raphson method. Given this value, we can use (2.51) to obtain

$$\beta = \log\left\{\frac{1}{\lambda}\left[\frac{m_1}{\Gamma(1+1/r)}\right]^{-r}\right\}. \tag{2.52}$$

As in Section 2.3.2, λ is a nuisance parameter and is given a proper uniform prior.

2.4.3 Example Analysis Using CMP Structure in the Weibull PHM

For an example analysis, we simulate 50 Weibull survival times, including 5 censored times, with parameter values: $r = 0.78$, $\beta = -0.35$, and nuisance parameter $\lambda = 0.08$. Given the relationships in (2.50) and (2.51), the simulated times have theoretical conditional means of $m_0 = 30$ and $m_1 = 47$. To gain information on β and r using the CMP method, we could attain expert opinion on m_0 and m_1. Suppose an expert provides the following modes and high values: $m_0^* = 35$, $m_0^U = 50$, $m_1^* = 53$, and $m_1^U = 70$. Taking $d = 3$ to be an unusually large value of the hazard ratio, the expert's information is represented in the following prior distributions:

$$\begin{aligned} m_0 &\sim \text{Gamma}(k_0 = 6.95, \theta_0 = 5.88), \\ m_1 &\sim \text{Gamma}(k_1 = 10.13, \theta_1 = 5.81), \end{aligned} \tag{2.53}$$

and

$$\lambda \sim \text{U}\left(0, \frac{\ln(2)}{3}\right). \tag{2.54}$$

To find the induced priors on the Weibull parameters, we simulate 100,000 values of m_0, m_1, and λ from their respective prior distributions in (2.53) and (2.54) in `R`. Specifying the functional relationships in `R` yields 100,000 values of r, β, and HR. The priors on the mean survival times and the resulting induced prior densities are shown in Figure 2.13 below.

The induced density plot on HR peaks around 0.78, which may be interpreted to mean that, most likely, the hazard rate of the experimental treatment group is 78% of the standard group. This implication should be verified with the expert. For the posterior analysis, we approximate the induced densities in (2.13) with independent priors

$$r \sim \text{Exp}(3.6, 0.49) \tag{2.55}$$

and

$$\beta \sim \text{N}(-0.22, 0.36^2). \tag{2.56}$$

where the prior on r is a shifted exponential as defined in Section 2.3.2.

Figure 2.13 also displays the approximate distributions on the induced priors. While the approximate distributions fit the densities well, there may be better-fitting approximations. Note, in particular, that the tails of β's induced priors are truncated by the approximate normal, which yields less variance. However, the t-distribution was not sufficiently concentrated for β, and the skew-normal truncated the tails, as well.

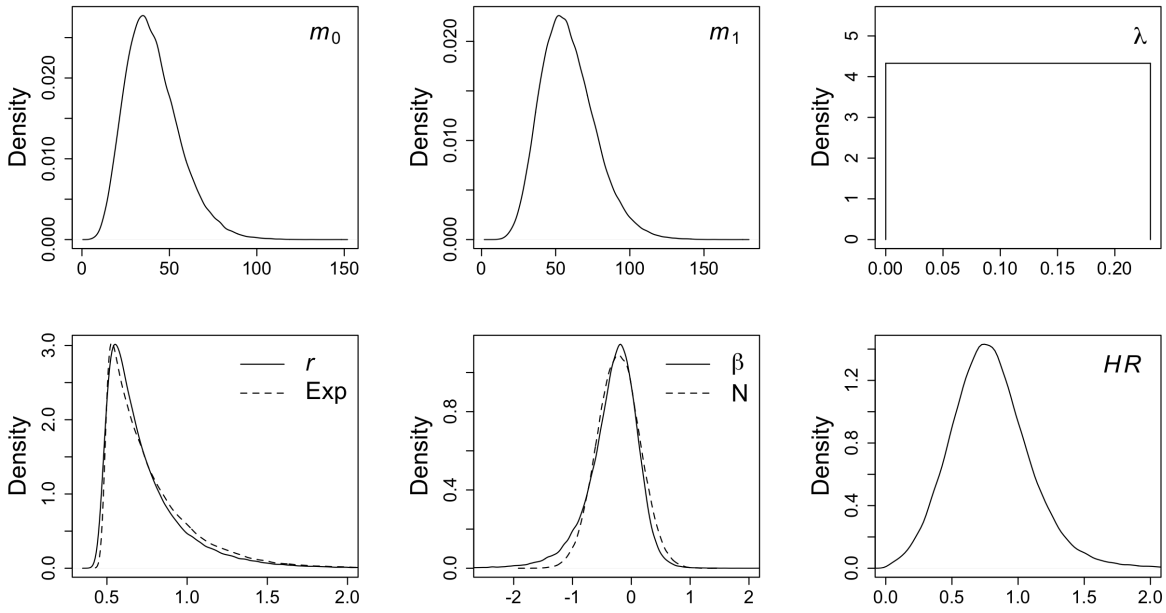


Figure 2.13: Priors on conditional means and nuisance parameter (top row) and resulting induced prior densities on r, β, and HR. Dotted lines represent approximate distributions.

As we noted above, the approximate independent priors in Figure 2.13 are not a perfect fit to the true induced priors. The primary purpose of the approximate independent priors is to facilitate feedback from an expert and aide prior-to-posterior sensitivity analyses. As we discussed in Sections 2.2 and 2.3.2, we recommend specifying the gamma priors on t_{m_0} and t_{m_1} and the functional relationships in (2.44) to implicitly induce the priors in OpenBUGS. This method will automatically capture any dependence, given their function relationships in (2.24) and (2.25), between the parameters.

The posterior summaries were computed using both the approximate independent priors and implicitly inducing the priors in OpenBUGS, both using a chain-length of 200,000 with 40,000 burn-in iterations. There were no convergence issues. Results are in Table 2.2 below. The MLE's were found in `R` by maximizing the Cox partial likelihood and the standard errors were computed using the delta method through the package `SurvRegCensCov`. In this example, the Bayesian posterior means using both prior structures for β are closer to the true value than the MLE. This is unsurprising given the use of informative priors. The MLE for r is identical to the posterior mean on r given the implicitly induced prior structure. However, in this one-off example, the MLE for λ is closer to the true value than both posterior means.

Table 2.3: Posterior results and frequentist estimates for Weibull PHM parameters using CMP prior structure

	Truth	Implicit Priors Model			Independent Priors Model			Frequentist Estimates	
		Mean	2.5%	97.5%	Mean	2.5%	97.5%	MLE	SE
λ	0.08	0.073	0.034	0.132	0.100	0.044	0.182	0.082	0.032
r	0.78	0.729	0.572	0.899	0.680	0.536	0.845	0.729	0.086
β	-0.35	-0.527	-1.027	-0.027	-0.442	-0.895	0.006	-0.583	0.300

Figure 2.14 displays the prior and posterior densities on the Weibull parameters. Solid dots indicate the MLE's.

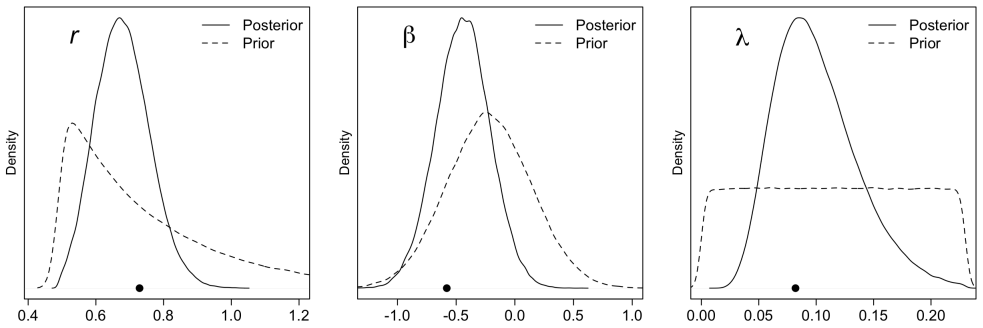


Figure 2.14: Prior and posterior densities on Weibull parameters resulting from the conditional means priors with solid dots at MLE's.

2.5 Models with Multiple Categorical Variables

In this section, we extend our methods to include multiple categorical variables with more than two levels. Consider a categorical covariate, x, with $\ell \geq 2$ levels. We can code x as $\ell - 1$ dichotomous variates, $z_1, \ldots, z_{\ell-1}$. (See, for example, Chapter eight of [33]). To be precise, suppose we have h multi-level covariates, $x_1, \ldots, x_h$, where x_i has ℓ_i levels. We convert x_i into $\ell_i - 1$ binary variates, $z_{i_1}, \ldots, z_{i_{\ell_i - 1}}$. We gather these h sets of binary variates into a $p \times 1$ vector, $\mathbf{z}$, where $p = \sum_{i=1}^{h} \ell_i - h$. The PHM is then

$$h(t \mid \mathbf{z}) = h_0(t) \exp(\beta_1 z_1 + \cdots + \beta_p z_p). \tag{2.57}$$

Specifying an exponential baseline hazard rate yields the exponential PHM given a vector of binary covariates

$$h(t \mid \mathbf{z}) = \gamma \exp(\beta_1 z_1 + \cdots + \beta_p z_p). \tag{2.58}$$

Similarly, we can specify a Weibull baseline hazard results in the Weibulll PHM given multiple binary covariates:

$$h(t \mid \mathbf{z}) = \lambda r t^{r-1} \exp(\beta_1 z_1 + \cdots + \beta_p z_p). \tag{2.59}$$

Our methodology can be readily extended to the PHM in (2.57) with multiple binary predictors (resulting from multiple categorical predictors). We outline the general algorithm for elicitation below.

Recall that we wish to induce a prior structure on the PHM parameters from expert information on median or mean survival times. To obtain information on the median/mean survival times, we construct $p+1$ variable configurations as shown in Table 2.4. The first configuration is a vector of 0's giving all covariates a zero value. From there, for $k = 1, \ldots p$, the $(k+1)^{th}$ configuration vector gives all covariates a zero value except for z_k which equals 1. Note that the configurations are hypothetical, for use only in the expert elicitation.

Table 2.4: Data partition with corresponding probabilities.

Configuration	z_1	z_2	z_3	z_4		z_p
1	0	0	0	0	...	0
2	1	0	0	0	...	0
3	0	1	0	0	...	0
4	0	0	1	0	...	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$p+1$	0	0	0	0	...	1

Then, we have $p+1$ median survival times, denoted $t_{m_0}, \ldots, t_{m_p}$, or mean survival times, $m_0, \ldots, m_p$, corresponding the the $p+1$ configurations. These are shown in Table 2.5.

Table 2.5: Data partition with corresponding probabilities for the exponential (Exp) and Weibull PHM's.

Configuration:		1	$k+1$
Exp	Median	$t_{m_0} = \log(2)/\gamma$	$t_{m_k} = \log(2)/[\gamma \exp(\beta_k)]$
	Mean	$m_0 = \gamma^{-1}$	$m_k = [\gamma \exp(\beta_k)]^{-1}$
Weibull	Median	$t_{m_0} = [\log(2)/\lambda]^{1/r}$	$t_{m_k} = \{\log(2)/[\lambda \exp(\beta_k)]\}^{1/r}$
	Mean	$m_0 = \lambda^{-1/r}\Gamma(1+1/r)$	$m_k = [\lambda \exp(\beta_k)]^{-1/r}\Gamma(1+1/r)$

The median and mean times in Table 2.5 were found via

$$S(t_{m_k} \mid \text{Configuration} = k+1) = 0.5, \quad k = 0, \ldots, p; \tag{2.60}$$

and

$$\text{E}(t \mid \text{Configuration} = k+1) = m_k, \quad k = 0, \ldots, p \tag{2.61}$$

where $S(t)$ and $\text{E}(t)$ are the survival function and the expected value, respectively, found for the appropriate distribution.

We induce priors on the PHM parameters in the following way. As before, we elicit a mode and upper percentile for each t_{m_k} (or m_k) for $k = 0, \ldots, p$. Gamma priors are constructed on the median (or mean) survival times representing this expert information.

For the exponential PHM, a prior on γ is induced through Configuration 1 (see Table 2.4) either through the median or mean survival time. Then, for $k = 1, \ldots, p$, via Configuration $k+1$, priors are induced on β_k through the induced prior on γ and the prior on t_{m_k} (or m_k).

For the Weibull PHM, we construct a uniform prior for λ by choosing d to be an unusually large value of HR following Section 2.3.2. However, here, there are multiple HR's to consider. Therefore, special care must be taken. If the data suggests that one or more of the true HR's is small, then the posterior density on λ will pile up at the upper bound. In this case, d should be decreased. Once a prior is given to λ, Configuration 1 induces a prior on r through expert information on t_{m_0} (or m_0). For $k = 1, \ldots, p$, via Configuration $k+1$, a prior is induced on each β_k through the induced prior on r and the prior on t_{m_k} (or m_k).

2.5.1 Exponential PHM with Multiple Categorical Variables

Consider an exponential PHM to model survival times t of patients with a progressive disease. Let x be a dichotomous variable denoting gender and z be a three-level categorical

variable denoting three stages of the disease.

$$h(t \mid x, z_1, z_2) = \gamma \text{exp}\left(\beta_1 x + \beta_2 z_2 + \beta_3 z_3\right) \tag{2.62}$$

where z_2 is 1 if the patient is in stage 2 of the disease and 0 otherwise; and z_3 is 1 if the patient is in stage 3 of the disease and 0 otherwise.

Because we have four parameters to model, we will have four median survival times about which to consult an expert. To obtain information on the median survival times, we construct four variable configurations. In particular, we have median survival times, denoted $t_{m_0}, \ldots, t_{m_3}$, shown in Table 2.6.

Table 2.6: Design configurations corresponding to particular scenarios for elicitation.

Configuration	x	z_2	z_3	Scenario	Median
1	0	0	0	Female, Stage 1	$t_{m_0} = \log(2)/\gamma$
2	1	0	0	Male, Stage 1	$t_{m_1} = \log(2)/[\gamma \exp(\beta_1)]$
3	0	1	0	Female, Stage 2	$t_{m_2} = \log(2)/[\gamma \exp(\beta_2)]$
4	0	0	1	Female, Stage 3	$t_{m_3} = \log(2)/[\gamma \exp(\beta_3)]$

We induce priors on the parameters in the following way. As before, we elicit a mode and upper percentile for each median survival time. Gamma priors are constructed on the median survival times representing this expert information. A prior on γ is induced through Configuration 1. Priors are induced on β_k through the induced prior on γ and the prior on t_{m_k} via Configurations $k+1$, $k = 1, \ldots, 3$.

To this end, suppose an expert gives most likely values and high values shown in Table 2.7. As before, we treat the most likely values and high values as modes and 75^{th} percentiles, respectively, of gamma distributions. The resulting gamma priors are shown in Table 2.7.

Table 2.7: Elicitation configurations corresponding to hypothetical scenarios with resulting priors on median times.

Configuration	Scenario	Mode	High Value	Prior
1	Female, Stage 1	65	92	$t_{m_0} \sim \text{Gamma}(7.23, 10.43)$
2	Male, Stage 1	60	87	$t_{m_1} \sim \text{Gamma}(6.55, 10.80)$
3	Female, Stage 2	45	75	$t_{m_2} \sim \text{Gamma}(4.22, 13.97)$
4	Female, Stage 3	40	70	$t_{m_3} \sim \text{Gamma}(3.75, 14.52)$

To examine the induced priors, we sample 100,000 values from the gamma priors on the median survival time. We specify in R the functional relationships between the median times and the PHM parameters, as in equations (2.24) and (2.25). This results in 100,000 values sampled from the induced priors on the PHM parameters. The priors are shown in Figure 2.15.

For an example analysis, we simulate 75 survival times, including 5 censored times, from the exponential PHM in (2.62). The x values are generated from a binomial distribution with rate 0.5, and z is constructed so that a third of the patients are in each of the three stages. The simulated data has the following true median times: $t_{m_0} = 75, t_{m_1} = 71, t_{m_2} = 53$, and $t_{m_3} = 50$. This yields the true PHM parameters: $\gamma = 0.009, \beta_1 = 0.06, \beta_2 = 0.35$, and $\beta_3 = 0.41$.

For analysis in OpenBUGS, we specify the gamma priors on the median survival times in the OpenBUGS model along with their functional relationships to the parameters of interest. This induces priors on γ and $\boldsymbol{\beta}$ and accounts for the dependence between them. We use a chain length of 500,000 after a burn-in of 100,000 iterations. There were no convergence issues. The posterior mean, 2.5th, and 97.5th posterior percentiles are shown in Table 2.8. Also, the frequentist estimates are shown, including the MLE, found in R by maximizing the Cox partial likelihood, and the standard error (S.E.) approximated by the Delta Method. The true values are within their respective 95% credible intervals and within one standard error of their MLE's. For all but one parameter, the posterior means are closer to the true values than the MLE's. We expect this because of the addition of the expert's

opinion in the Bayesian model. The posteriors with their corresponding priors are shown in Figure 2.16.

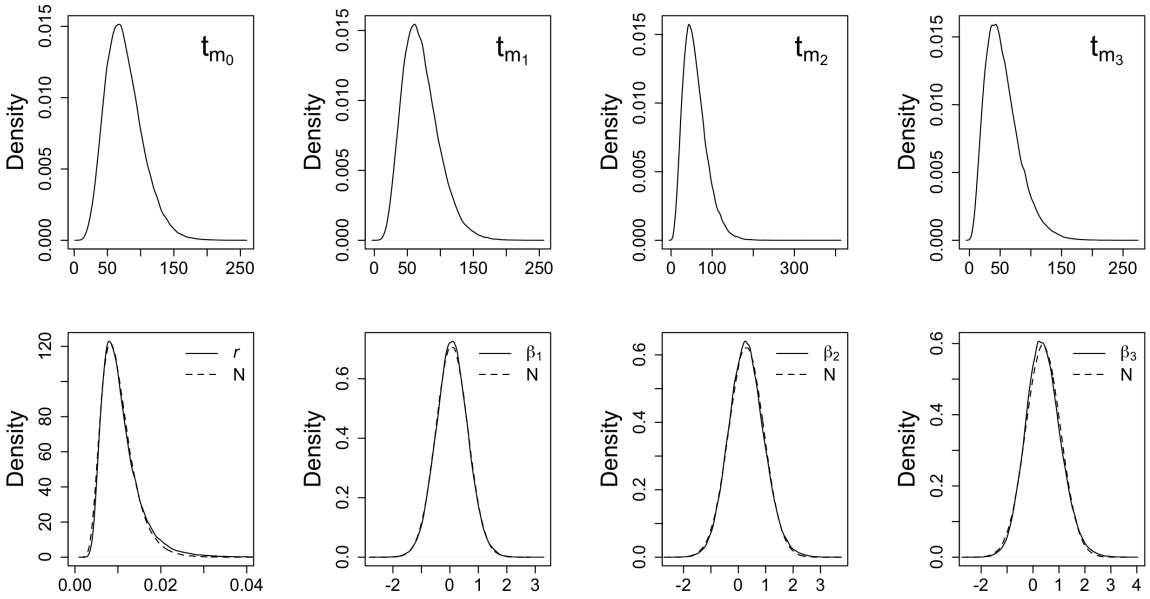


Figure 2.15: Priors on median survival times (top) and induced priors on PHM parameters (bottom). Dotted lines represent approximate independent densities.

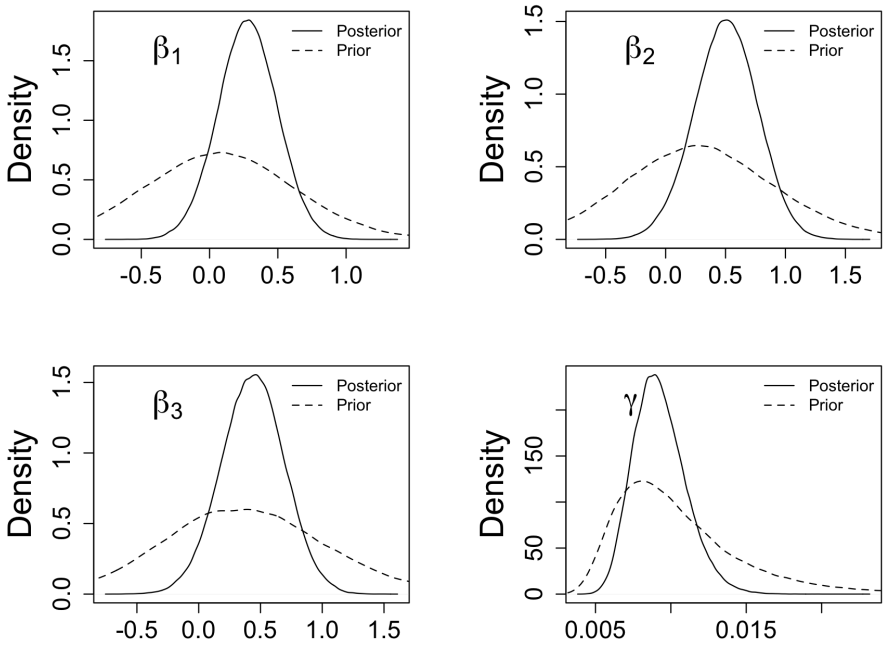


Figure 2.16: Prior to posterior plots for the exponential PHM with multiple covariates.

Alternatively, we can approximate independent prior distributions for the PHM parameters. To illustrate, we fit the induced priors with normal distributions, shown as dotted lines in Figure 2.15. The approximate normal priors are appear to be good fits. The independent priors on the parameters are

$$\begin{aligned} \gamma &\sim \text{LN}(-4.66, 0.37^2), \\ \beta_1 &\sim \text{N}(0.07, 0.56^2), \\ \beta_2 &\sim \text{N}(0.29, 0.64^2), \end{aligned} \tag{2.63}$$

and

$$\beta_3 \sim \text{N}(0.39, 0.67^2).$$

Specifying these independent priors for the analysis results in posterior summaries shown in Table 2.8 along with the posterior summaries resulting from use of the implicitly induced priors.

Table 2.8: Posterior results and frequentist estimates for exponential PHM parameters with multiple covariates.

	Truth	Implicit Priors Model			Independent Priors Model			Frequentist Estimates	
		Mean	2.5%	97.5%	Mean	2.5%	97.5%	MLE	SE
γ	0.009	0.009	0.006	0.013	0.01	0.007	0.014	0.0096	0.002
β_1	0.06	0.279	−0.148	0.705	0.26	−0.172	0.694	0.294	0.244
β_2	0.35	0.503	−0.021	1.102	0.47	0.046	0.972	0.506	0.302
β_3	0.41	0.437	−0.068	0.935	0.40	−0.104	0.230	0.411	0.295

2.6 Discussion

In this chapter, we have presented a prior elicitation method for use in parametric proportional hazards models. We focused on the exponential and the Weibull PHM's and elicited expert opinion on median and mean survival times. Such elicitation is desirable because experts are often more familiar and comfortable with observables such as the survival time summaries. We also discussed the Weibull PHM parameterization which gave several additional ways of gathering expert information on observables.

In our examples, we assumed the experts gave accurate information. The prior formulation on the medians and means induce priors that are consistent with the truth on the parameters of interest, and our model does a good job of estimating the true parameters *a posteriori*. Also in some data sets, our model better estimates the true parameters than the frequentist MLE's reflecting our use of informative priors. Indeed, the induced priors are typically informative and have a relatively substantial affect on the posterior inferences. Therefore, necessary caution should be taken when applying the priors, such as prior-to-posterior analyses and checking prior implications with the expert.

In the first sections, we focused on eliciting opinion on median survival times. We then presented methodology for the means, known as conditional means priors. Both techniques produced posterior results close to the true values. We note that in the data simulation, median and means often differed because of the skewness of the distribution. Also, medians typically characterized the data better than the means, which is expected due to the skewed nature of the Weibull and exponential distributions. For this reason, it is important for the expert to have a good understanding of the inherent difference between medians and means.

It is common in the literature to give the parameters independent priors although there is a functional relationship between them. Ibraham and Chen [26] and Christensen et al. [10] assume the same *a priori* independence in the parameters. Others that treat Weibull PHM parameters as independent include: Albert [3], the WinBUGS help manual [50], Hamanda et al. [25], and Canavos and Tsokos [5]. Papers treating these priors as dependent include Soland [48], Erto [22], and Rinne [46].

The priors can be induced in OpenBUGS implicitly by specifying the relationships between the parameters and median survival times. This automatically accounts for the dependence between the parameters. So far, we have seen that the resulting posteriors differ slightly from the posteriors given independent approximate priors in the Weibull PHM.

CHAPTER THREE

Operational Assessments for Determining a Power Prior

Often in planning experiments, historical data is available from previous studies that assess the same quantity of interest. Such data can be used in the construction of prior distributions but care must be taken to gauge similarity to "current" data and avoid dominance of the "new" likelihood. Power priors are one way to incorporate previous research into current trials with these goals in mind. The power prior concept was introduced by Chen et al. [9]. The review paper by Chen and Ibrahim [26] details the background of power priors. See also Chen and Ibrahim [8].

In Section 3.1, we present the general form of the power prior. In Section 3.2, we discuss operational assessment of power priors using prior effective sample size (ESS). In Section 3.3, motivated by FDA guidelines, we suggest another way of selecting power priors. Corresponding examples are presented in subsequent subsections. We consider the use of multiple studies in Section 3.4. We briefly consider power prior specification with multiple experts in Section 3.5. The chapter concludes with a discussion in Section 3.6.

3.1 General Form of the Power Prior

Suppose we have data from a current experiment that we would like to analyze using a Bayesian model. Further assume that we have historical data from a previous, similar experiment that we would like to incorporate into our current analysis. The idea behind power priors is to form a posterior from the historical likelihood and a vague prior, and to use this posterior as the prior for the current study. Doing so without attenuating the variance of the historical data can potentially overwhelm the current data. Therefore, the

power prior formulation raises the historical likelihood to a weighting exponent called the power parameter.

Formally, suppose there is a historical study with sample size n_0, data vector $\mathbf{y}_0$, covariate matrix $\mathbf{X}_0$, and likelihood $\ell(\boldsymbol{\theta}|D_0)$, where $D_0 = (n_0, \mathbf{y}_0, \mathbf{X}_0)$. Let π_0 be a prior on $\boldsymbol{\theta}$ which is independent of D_0 and D. This is called the initial prior for $\boldsymbol{\theta}$. The conditional power prior distribution for $\boldsymbol{\theta}$ is defined as

$$\begin{aligned}\pi(\boldsymbol{\theta}|D_0, a_0) &= \frac{[\ell(\boldsymbol{\theta}|D_0)]^{a_0}\pi_0(\boldsymbol{\theta})}{\int_\Theta \ell(\boldsymbol{\theta}|D_0)^{a_0}\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}} \\ &\equiv C(a_0)[\ell(\boldsymbol{\theta}|D_0)]^{a_0}\pi_0(\boldsymbol{\theta}) \end{aligned} \tag{3.1}$$

where we have written the normalizing constant as

$$C(a_0) = \left[\int_\Theta \ell(\boldsymbol{\theta} \mid D_0)^{a_0}\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}\right]^{-1}. \tag{3.2}$$

The power prior will be proper so long as the integral in the normalizing constant is finite [39]. The power prior is typically written as

$$\pi(\boldsymbol{\theta}|D_0, a_0) \propto [\ell(\boldsymbol{\theta} \mid D_0)]^{a_0}\,\pi_0(\boldsymbol{\theta}). \tag{3.3}$$

We refer to $a_0 \in [0, 1]$ as the power parameter.

There are two alternative forms of the power prior, the joint and the modified power priors. Chen and Ibrahim [26], [8], define the joint power prior for $\boldsymbol{\theta}$ and a_0 as

$$\begin{aligned}\pi(\boldsymbol{\theta}, a_0 \mid D_0) &= \frac{[\ell(\boldsymbol{\theta} \mid D_0)]^{a_0}\pi_0(\boldsymbol{\theta})\pi(a_0)}{\int_0^1\int_\Theta [\ell(\boldsymbol{\theta} \mid D_0)]^{a_0}\pi_0(\boldsymbol{\theta})\pi(a_0)d\theta da_0} \\ &\propto [\ell(\boldsymbol{\theta} \mid D_0)]^{a_0}\pi_0(\boldsymbol{\theta})\pi(a_0).\end{aligned} \tag{3.4}$$

Problems with the form in (3.4) have been noted in several papers, starting with Duan and Smith in 2006 [20], [19], [40], [39]. Nueuschwander et al. [40] present the modified power prior to remedy the noted problems with the joint power prior. The modified power prior is

$$\pi(\boldsymbol{\theta}, a_0 \mid D_0) = C(a_0)[\ell(\boldsymbol{\theta} \mid D_0)]^{a_0}\pi_0(\boldsymbol{\theta})\pi(a_0) \tag{3.5}$$

where $C(a_0)$ is the normalizing constant in (3.2). Both the joint and modified power priors treat a_0 as random which we do not recommend for reasons we note below.

The power parameter a_0 controls the effect of the historical data in the current model. When $a_0 = 0$, no historical data is used in the current model. If we take $a_0 = 1$ then (3.1) reduces to Bayes' theorem, and the resulting posterior becomes the prior for the current study analysis. The power parameter controls the heaviness of the tails of the prior for $\boldsymbol{\theta}$, with smaller values yielding heavier tails. Essentially, the power parameter is a weight for the historical data relative to the likelihood. More discounting of the historical model is necessary as the current and historical studies become less commensurable.

As we will see, specification of the power parameter is critical to the model and has substantial implications. However, the literature on the topic is rather limited. Ibrahim & Chen (2000) [26] use random a_0 in models including the proportional hazards model, the generalized linear model and the cure rate model. Duan et al. (2008) [21] use a compatibility statistic to quantify a_0's effect. Reitbergen et al. (2011) [45] present a sensitivity analysis for a_0 and elicit ranks from an expert to select a fixed a_0.

While some specify a prior for a_0, doing so can be problematic. The hope is that the data will update the power parameter's prior with information on the heterogeneity between the current and historical trial, but this typically does not occur. Results found by Nuenschwander et al. [40] imply that the heterogeneity of the data cannot be assed precisely, especially for small sample sizes. With this in mind, we focus our methods on selection of a fixed a_0.

3.1.1 Example

To illustrate the general construction of the power prior, we generate n_0 exponential survival times $\mathbf{t}_0 = (t_{01}, \ldots, t_{0n_0})$ from an exponential proportional hazards model dependent on a binary covariate $\mathbf{z}_0$. That is, let $z_{0i} \in \{0, 1\}$ indicate whether (1) or not (0) the i^{th} patient is in the experimental treatment group. This will serve as the historical data for our

example. Thus, for $i = 1, \ldots, n_0$,

$$t_{0i} \sim \text{exponential}\left(\gamma \exp(\beta z_{0i})\right).$$

where γ and β are unknown (see Section 2.1.2). Then, the likelihood for β and γ is

$$\begin{aligned} \ell(\beta, \gamma \mid \mathbf{t}_0, \mathbf{z}_0, \boldsymbol{\delta}_0) = \prod_{i=1}^{n_0} &\{\gamma \text{exp}(\beta z_{0i}) \text{exp}\left[-t_{0i}\gamma \text{exp}(\beta z_{0i})\right]\}^{\delta_{0i}} \\ &\times \{\text{exp}\left[-t_{0i}\gamma \text{exp}(\beta z_{0i})\right]\}^{1-\delta_{0i}} \end{aligned} \tag{3.6}$$

where $\boldsymbol{\delta}$ is an $n_0 \times 1$ vector of censoring indicators, with $\delta_{0i} = 1$ if the i^{th} historical death time is observed and zero otherwise.

We use independent non-informative initial prior densities on β and γ. These are diffuse normal and exponential initial priors densitites, denoted by $\phi_0(\mu, \sigma^2)$ and $\psi_0(\nu)$, respectively, on β and γ. Let the historical data be $D_0 \equiv (\mathbf{t}_0, \mathbf{z}_0, \boldsymbol{\delta_0})$. Then the power prior is

$$\begin{aligned} \pi(\beta, \gamma \mid D_0, a_0) = \prod_{i=1}^{n_0} &\{\gamma \text{exp}(\beta z_{0i}) \text{exp}\left[-t_{0i}\gamma \text{exp}(\beta z_{0i})\right]\}^{a_0\delta_{0i}} \\ &\times \{\text{exp}\left[-t_{0i}\gamma \text{exp}(\beta z_{0i})\right]\}^{a_0(1-\delta_{0i})} \phi_0(\mu, \sigma^2)\psi_0(\nu). \end{aligned} \tag{3.7}$$

The density of the power prior is found by sampling from the posterior in (3.7) using OpenBUGS. Power priors on β and γ are shown in Figure 3.1. To illustrate the effect of the power parameter, the graphs show the increase in variance from $a_0 = 1$ to $a_0 = 0.2$. Clearly, the choice of a_0 is critical. We now turn to methods for its selection.

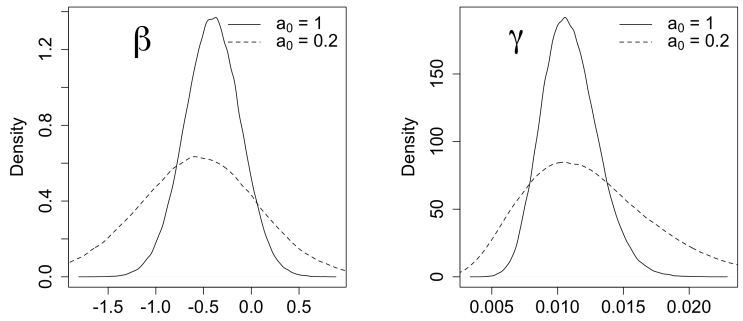


Figure 3.1: Power priors for β and γ with decreasing power parameters.

3.2 Using Equivalent Sample Size to Determine Power Parameters

Selection of a fixed value of the power parameter, a_0, requires an operational assessment of the implications of that choice. In this section we suggest criteria for assessing the posterior impact of a value of a_0 based on prior equivalent sample size (ESS), using the methods of Morita et al. (2008) [36]. We present three examples using logistic regression models. In Section 3.2.2 we consider a_0 specification in power priors on logistic regression coefficients. We extend this example in Section 3.2.3 to the logistic regression probabilities. Lastly, in Section 3.2.4 we use a logistic regression model with multiple categorical variables to illustrate the use of the prior ESS in a_0 selection.

The prior effective sample size is a useful tool for determining the effect of the prior relative to the likelihood. The ESS reflects the amount of information contained in the prior. It is easily understood by non-statisticians. In effect, the ESS is that number of observations that would need to be added to the likelihood in order to yield a comparable analysis using a relatively non-informative prior.

Morita et al. [36] discuss several points of application for the prior ESS. First, values of the ESS can be used as feedback when eliciting expert information. The expert can verify or change her opinion based on the resulting prior ESS. Alternatively, a small ESS can be used to justify a choice of a relatively non-informative prior. More details and motivation can be found in Section 4 of [36]. If the prior ESS is larger than the sample size of our current study, the information in the prior could overwhelm the current study information. We propose selecting a_0 so as to avoid this. As Morita et al. (2008) [36] note,

> For example, when fitting a Bayesian model to a data set of 10 observations, an a priori ESS of 1 is reasonable, whereas a prior ESS of 20 implies that the prior, rather than the data, dominates posterior inferences. If the prior is elicited from a domain expert, then an informative prior is desirable ... In contrast, if the prior is only a technically convenient ad hoc choice, as is often the case in practice, then understanding the ESS may prompt the investigator to reconsider the prior choice.

Morita el al. [36] explain the intuition behind the prior ESS using a Beta(a, b) prior. Note that a Beta(a, b) prior together with a binomial random variable y from a sample of size n forms a Beta$(a - y, b + n - y)$ posterior. So, the prior has been updated with n Bernoulli observations. Thus, any Beta(a, b) prior can be identified with a Beta$(c + y, d + m - y)$ posterior resulting from a diffuse Beta(c, d) prior and m Bernoulli observations. Setting $c + d = \epsilon$, for small ϵ will form a diffuse Beta(c, d) prior, and letting $c + d + m = a + b$ gives $m = a + b - \epsilon$.

Morita el al. [36] describe the prior ESS as follows. Suppose that we have a prior, $\pi(\boldsymbol{\theta})$, for a parameter $\boldsymbol{\theta} \in \mathbb{R}^p$, for which we wish to determine the ESS. To do so, we construct a vague, "ϵ-informative" prior $\pi_0(\boldsymbol{\theta})$ and, using some vector of observations, $\mathbf{y}$, of length m, we find the posterior $\pi_m(\boldsymbol{\theta} \mid \mathbf{y})$. Note that this vector, $\mathbf{y}$, is not used for posterior inferences but is hypothetical for use in determining m. Morita et al. [36] provide suggestions for $\pi_0(\boldsymbol{\theta})$. The prior ESS is the value of m that minimizes the distance between $\pi_m(\boldsymbol{\theta} \mid \mathbf{y})$ and $\pi(\boldsymbol{\theta})$. This distance is defined as the difference between the trace of the Fisher information matrix of $\pi(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}})$ and the expected information matrix of $\pi_m(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}_0, \mathbf{y})$, where $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}_0$ are vectors of hyperparameters. In the latter, the expectation is taken with respect to the marginal distribution of the data

The definition of the prior ESS allows for calculation of ESS's for sub-vectors of $\boldsymbol{\theta}$. For example, consider $\boldsymbol{\theta} = (\theta_0, \theta_1)$. As we shall see, we not only find a prior ESS for $\boldsymbol{\theta}$, but also individually for θ_0, and for θ_1. The latter are computed through use of their marginal prior distributions. However, Morita et al. [37] note that ESS(θ_0) + ESS(θ_1) usually does not equal the ESS(θ_0, θ_1). This is because θ_i typically has a different meaning in its marginal distribution than in its joint distribution.

3.2.1 Method for Assessing Power Parameter Choices using the Prior ESS

In this section, we narrow choices for a_0 by eliminating those that yield prior ESS's in excess of the proposed sample size for the current study. Such power parameters will

likely yield power priors that overwhelm information in the likelihood thereby dominating posterior inference. We present prior ESS's for the full power prior and for parameter sub-vectors, e.g. for $\boldsymbol{\theta}$ and for the component θ_i's. Note that, while a large prior ESS on, say, θ_0, suggests that the power prior contains a large amount of information on θ_0, it does not automatically rule out that power prior. However, the ESS's on parameter sub-vectors are still useful in our problem. If the scientist specifically desires little information on θ_0 in the prior, then the power parameter should be chosen so that the prior ESS on θ_0 relatively small. For example, in a regression setting with two coefficients, there may be no prior knowledge of the intercept but much information on the slope. Thus, the intercept would be given a diffuse prior with little information so that the ESS is relatively small, but the slope may have a fairly large prior ESS. We recommend a prior-to-posterior sensitivity analysis for large marginal ESS's. See Morita et al. [36] and [37] for more details on interpretation of prior ESS's on parameter sub-vectors.

To illustrate our proposed method for eliminating a_0 values based on prior ESS, we present a simple non-regression example using a power prior for binomial sampling. Consider historical data, D_0, consisting of binomial responses:

$$y_{0i} \sim \text{Bin}(n_0, \theta), i = 1, \ldots, N_0.$$

We use the initial prior $\theta \sim \text{U}(0, 1)$. Then the power prior is, for some power parameter $a_0 \in [0, 1]$,

$$\pi(\theta) \propto \left(\theta^{\sum y_{0i}}(1-\theta)^{\sum(n_0-y_{0i})}\right)^{a_0} I_{(0,1)}(\theta). \tag{3.8}$$

The right hand side of (3.8) is proportional to a beta distribution with shape parameters $\alpha_p \equiv a_0 \sum y_{0i} + 1$ and $\beta_p \equiv a_0 \sum(n_0 - y_{0i}) + 1$. Essentially, the power prior represents the information in $\alpha_p + \beta_p$ Bernoulli trials yielding α_p successes. Thus, the power prior in (3.8) has a prior ESS, m, of

$$\begin{aligned} m &= a_0 \sum y_{0i} + 1 + a_0 \sum(n_0 - y_{0i}) + 1 \\ &= a_0(N_0 n_0) + 2. \end{aligned} \tag{3.9}$$

Since m is a linear function of a_0 the latter is easily evaluated. To this end, the prior ESS should be compared to the sample size of the current data model. For example, if the historical data consisted of $N_0 = 25$ observations of binomial random variables based on $n_0 = 10$ Bernoulli trials, the prior ESS is

$$m(a_0) = 250a_0 + 2.$$

If the current data set consists of 100 Bernoulli observations, then we find the maximum plausible a_0 by setting $m(a_0) = 100$:

$$\begin{aligned} 100 &= 250(a_0) + 2 \\ \Leftrightarrow a_0 &= 0.392 \end{aligned} \tag{3.10}$$

The remaining choices for a_0 could be discussed with subject-matter experts.

Finally, the subject-matter expert can provide an assessment of the historical data by specifying an equivalent sample size. This can then be used to determine a_0. Clearly this approach can be problematic. We do not recommend this approach because of the obvious potential bias. It might be more reasonable for the expert to provide an upper bound on the ESS, perhaps as a fraction of the planned sample size. We consider this idea further in Section 3.2.3.

3.2.2 Power Prior on Regression Coefficients

We now consider a logistic regression with one dichotomous covariate and use the prior ESS to evaluate fixed power parameter choices. Here, the power prior is given to the vector of logistic regression coefficients. We use an `R` program to calculate the prior ESS.

Consider a historical logisitic regression model with one binary covariate, so that

$$y_{0i} \sim \text{Bernoulli}(p_i) \tag{3.11}$$

and

$$p_i = \left\{1 + \exp\left(-\beta_0 - \beta_1 x_{0i}\right)\right\}^{-1}.$$

The likehood function for $\boldsymbol{\beta} = (\beta_0, \beta_1)$ given the historical data, D_0, is

$$\begin{aligned}\ell(\boldsymbol{\beta} \mid D_0) &= \prod_{i=1}^{n_0} \left(\frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{y_{0i}} \left(1 - \frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{(1-y_{0i})} \\ &= \left(\frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{a_0 \sum_{i=1}^{n_0} y_{0i}} \left(1 - \frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{a_0 \sum_{i=1}^{n_0} (1-y_{0i})}.\end{aligned} \tag{3.12}$$

Following the development in Neelon & O'Malley [39], we use independent and identical diffuse normal initial prior densities, denoted by $\phi_0(\mu, \sigma^2)$, on β_0 and β_1. Then the power prior is [1]

$$\begin{aligned}\pi(\boldsymbol{\beta} \mid D_0, a_0) &\propto \prod_{i=1}^{n_0} \left(\frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{a_0 y_{0i}} \\ &\quad \times \left(1 - \frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{a_0(1-y_{0i})} \phi_0(\mu, \sigma^2)\phi_0(\mu, \sigma^2) \\ &= \left(\frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{a_0 \sum_{i=1}^{n_0} y_{0i}} \\ &\quad \times \left(1 - \frac{1}{1+\exp(-\beta_0 - \beta_1 x_{0i})}\right)^{a_0 \sum_{i=1}^{n_0} (1-y_{0i})} \phi_0(\mu, \sigma^2)\phi_0(\mu, \sigma^2).\end{aligned}$$

For this example we take $\mu \sim \text{N}(0, 1000)$ and $\sigma \sim$ Half-N$(0, 1)$, which is the positive half of a N$(0, 1)$ distribution. We find the power prior in (3.13) using Markov chain Monte Carlo (MCMC) methods in OpenBUGS. Prior ESS is used to judge the affect of a_0 on the power prior. To this end, we find the prior ESS for the vector $\boldsymbol{\beta}$ as well as β_0 and β_1 individually, using the `R` program `ESS_RegressionCalculator`. [2]

To illustrate, consider simulated logistic regression data with $\beta_0 = -1$ and $\beta_1 = 0.16$ and $n_0 = 1000$. The design is balanced, with 500 responses for $x_0 = 0$ and 500 for $x_0 = 1$.

[1] Note that here, and in regression models throughout the chapter, we assume that the historical and current model share identical covariates. Neelon & O'Malley [39] discuss relaxing this assumption. They instead assume exchangeable covariates, giving them different parameters but the same prior distribution.

[2] Created to correspond with Morita et al. (2008) [36], this program can be downloaded from the M.D. Anderson website at https://biostatistics.mdanderson.org/SoftwareDownload/SiteAux/Tags.html. (See `ESS_Regression` under the `Bayesian` program section heading. `ESS_RegressionCalculator.R` requires that the prior distributions on the parameters of interest be mutually independent and either normal or gamma distributions.

We use $a_0 = 0.85$. The resulting power priors, along with their normal approximations, are shown in Figure 3.2. The normal approximations are $\text{N}(-0.95, 0.11^2)$ and $\text{N}(0.14, 0.15^2)$, respectively. Specifying these distributions in `ESS_RegressionCalculator` yields prior ESS's of 443, 688, and 442 for β_0, β_1, and $\boldsymbol{\beta}$, respectively.

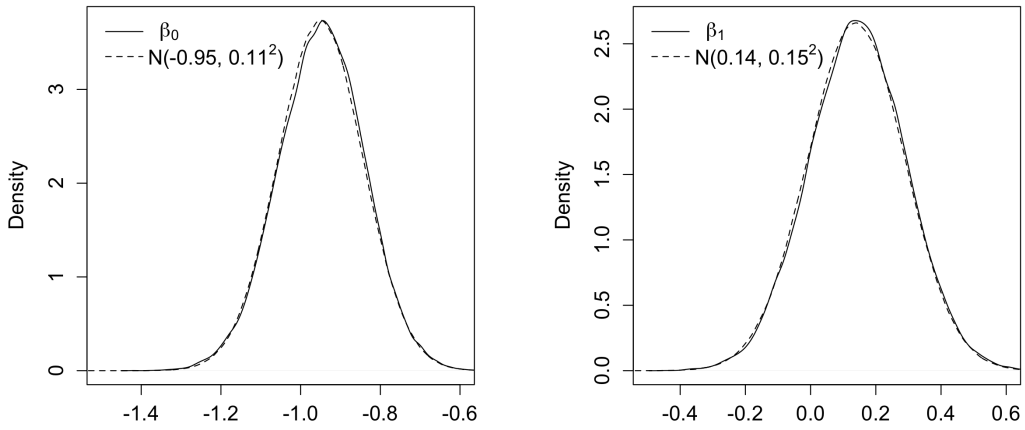


Figure 3.2: Power priors and their normal approximations on β_0 and β_1.

To assess power parameter choices, we repeat the computation above for varying values of a_0. Morita et al. [36] note the importance of specifying a prior ESS for the parameter vector as well as its components. Figure 3.3 shows the prior ESS's for β_0, β_1, and the vector $\boldsymbol{\beta}$ as a function of a_0.

Now, suppose we are planning a study with 300 data points, and we want to use the historical data set described above, which has a sample size of $n_0 = 1000$. The implications of selecting a_0 can now be seen in terms of prior ESS. For example, selecting $a_0 = 0.6$ yields a prior ESS of approximately 292, 485, and 311 for estimation of β_0, β_1, and $\boldsymbol{\beta}$, respectively.

In the drug-development context, choosing between the remaining a_0 values could be narrowed further by eliminating values of a_0 which are counter to FDA guidelines; see Section 3.3 for more on this issue.

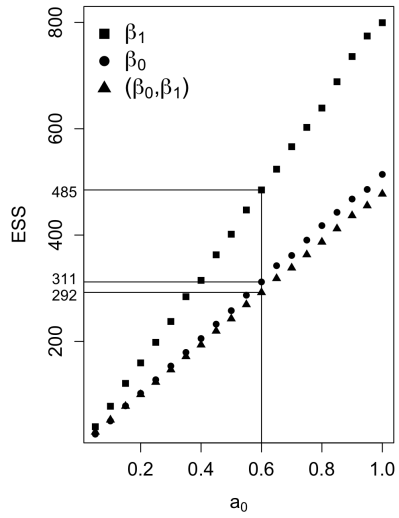


Figure 3.3: Prior ESS for regression coefficients given values of a_0.

Alternatively, Figure 3.3 can be used in conjunction with an expert-elicited upper bound on the ESS value, although this invites potential bias. To illustrate, suppose an expert gives information based on experience with similar models containing both β_0 and β_1. Then interest is in the prior ESS on the vector (β_0, β_1). (See Morital et al. [37] and Example 7 in [36] for more details.) Say that, based on the commensurability between the historical and current studies, the expert recommends an upper bound on the prior ESS of 200. Then, power parameter between 0.4 and 0.45 would be appropriate. The potential for bias is obvious here. Consequent posterior quantities, such as the probability of study success in drug development, should be carefully considered.

3.2.3 Power Prior on Regression Probabilities

In Section 3.2.2 above, we analyzed the choice of the fixed power parameter, a_0, based on methodology beginning with the power prior on the regression coefficients. Here, we present another approach which begins with the power prior on the regression probabilities, p_i in (3.11). In this approach, the prior ESS is determined analytically without a need for the empirical computations. Also, experts may be more familiar with the regression probabilites.

Assume the same model (3.11) above, and consider the success probabilities. Since our model contains only one binary covariate x_0, we have $p_i = p^0$ or $p_i = p^1$ given $x_0 = 0$ and $x_0 = 1$, respectively. We use superscripts here to avoid confusion with the observation indices, e.g. the first observation has success probability p_1 which could equal p^0. Formally, the probabilities are defined as

$$p^0 = \frac{1}{1 + \exp(-\beta_0)} \tag{3.13}$$

and

$$p^1 = \frac{1}{1 + \exp(-\beta_0 - \beta_1)}. \tag{3.14}$$

Let n_0^0 and n_0^1 be the number of observations such that $x_0 = 0$ and $x_0 = 1$, respectively. Then $n_0 = n_0^0 + n_0^1$. We partition the data into n_0^0 observations, $\mathbf{y}_0^0$, given $x_0 = 0$ and n_0^1 observations, $\mathbf{y}_0^1$, given $x_0 = 1$. Then, the likelihood of $\mathbf{p} \equiv (p^0, p^1)$, given the historical data, $D_0 \equiv (\mathbf{y}_0^0, \mathbf{y}_0^1)$ is

$$\ell\left(\mathbf{p} \mid D_0\right) = \prod_{i=1}^{n_0^0} \left(p^0\right)^{y_{0i}^0} \left(1 - p^0\right)^{(1-y_{0i}^0)} \prod_{i=1}^{n_0^1} \left(p^1\right)^{y_{0i}^1} \left(1 - p^1\right)^{(1-y_{0i}^1)}. \tag{3.15}$$

This likelihood is the product of independent Beta$(\sum y_{0i}^0 + 1, \sum(1 - y_{0i}^0) + 1)$ and Beta$(\sum y_{0i}^1 + 1, \sum(1 - y_{0i}^1) + 1)$ densities. [3] Consider the power prior with initial priors $\pi_0(p^j)$ for $j = 0, 1$ taken to be U$(0, 1)$, and with the likelihood in (3.15). Then the power prior on $\mathbf{p}$ is

$$\begin{aligned}
\pi\left(\mathbf{p} \mid \mathbf{D}_0, a_0\right) &\propto \left\{ \prod_{i=1}^{n_0^0} \left(p^0\right)^{y_{0i}^0} \left(1 - p^0\right)^{(1-y_{0i}^0)} \prod_{i=1}^{n_0^1} \left(p^1\right)^{y_{0i}^1} \left(1 - p^1\right)^{(1-y_{0i}^1)} \right\}^{a_0} \prod_{j=1}^{2} \mathrm{I}_{[0,1]}(p^j) \\
&= \left(p^0\right)^{a_0 \sum_{i=1}^{n_0^0} y_{0i}^0} \left(1 - p^0\right)^{a_0 \sum_{i=1}^{n_0^0} (1-y_{0i}^0)} \mathrm{I}_{[0,1]}(p^0) \\
&\times \left(p^1\right)^{a_0 \sum_{i=1}^{n_0^1} y_{0i}^1} \left(1 - p^1\right)^{a_0 \sum_{i=1}^{n_0^1} (1-y_{0i}^1)} \mathrm{I}_{[0,1]}(p^1).
\end{aligned} \tag{3.16}$$

[3] This is a bivariate beta distribution if $n_0^0 = n_0^1$. Specifically, let $n_0^0 = n_0^1 = n^*$, then (3.15) is a bivariate beta distribution with PDF $f(p^0, p^1) \propto \left(p^0\right)^{a-1} \left(p^1\right)^{b-1} \left(1 - p^0\right)^{b+c-1} \left(1 - p^1\right)^{a+c-1}$ where $a = \sum y_{0i}^0 + 1$, $b = \sum y_{0i}^1 + 1$ and $c = n^* - \sum y_{0i}^0 - \sum y_{0i}^1$. See [42].

Morita et al. [36] give a general form for calculating the prior ESS of a power prior for a beta/binomial model, as we have here. Consider the power power $\pi\left(\mathbf{p} \mid D_0, a_0\right)$ on the vector of probabilities $\mathbf{p}$. The authors give the prior ESS for $\pi\left(\mathbf{p} \mid D_0, a_0\right)$ as

$$m_{\mathbf{p}} = a_0 \times \text{ESS}\left\{\ell\left(\mathbf{p} \mid D_0\right)\right\} + \text{ESS}\left\{\pi_0(\mathbf{p})\right\}. \tag{3.17}$$

That is, the ESS of the power prior is a_0 times the ESS of the historical likelihood plus the ESS of the inital prior.[4] We will see that our marginal priors on sub-vectors of $\mathbf{p}$ do, in fact, have ESS's of this form.

The power prior (3.16) implies that we have independent beta priors on p^0 and p^1:

$$p^0 \mid \mathbf{y}_0^0, a_0 \sim \text{Beta}\left(a_0 \sum y_{0i}^0 + 1, a_0 \sum (1 - y_{0i}^0) + 1\right) \tag{3.18}$$

and

$$p^1 \mid \mathbf{y}_0^1, a_0 \sim \text{Beta}\left(a_0 \sum y_{0i}^1 + 1, a_0 \sum (1 - y_{0i}^1) + 1\right). \tag{3.19}$$

The beta distributions above allow for a straight forward interpretation of the effect of a_0 through the use of prior ESS. The prior ESS for a Beta(a, b) prior distribution with a binomial likelihood is $m = a + b$. Therefore, the prior ESS for p^0 and p^1 is, respectively,

$$\begin{aligned} m^0 &= a_0 \sum y_{0i}^0 + 1 + a_0 \sum (1 - y_{0i}^0) + 1 \\ &= a_0 n_0^0 + 2 \end{aligned} \tag{3.20}$$

and

$$\begin{aligned} m^1 &= a_0 \sum y_{0i}^1 + 1 + a_0 \sum (1 - y_{0i}^1) + 1 \\ &= a_0 n_0^1 + 2. \end{aligned} \tag{3.21}$$

[4] There is an apparent error in Morita et al. They refer to a power prior $p(\theta) \propto \left[\theta^3(1-\theta)^7\right]^{a_0} \theta(1-\theta)$ arising from a binomial likelihood and a Beta(1,1) initial prior. First, with a Beta(1,1) initial prior, the power prior should be $p(\theta) \propto \left[\theta^3(1-\theta)^7\right]^{a_0}$. The likelihood is Beta(4,8), so it has an ESS of 12. However, they refer to the historical likelihood having an ESS of 10, so that the full prior ESS is $10a_0 + 2$ where the $+2$ comes from the ESS of the Beta(1,1) initial prior. The ESS of $10a_0 + 2$ is correct, but it's not a_0 times the *ESS of the historical likelihood as a function of* θ plus the ESS of the initial prior. For the beta/binomial case, the correction could read "a_0 times (the ESS of the historical likelihood as a function of θ minus 2) plus the ESS of the initial prior" (or simply subtract $2a_0$ from their definition), which is what we will use in this section.

Recall Morita et al.'s prior ESS for **p** in (3.17). We deduce from (3.15) that the marginal likelihoods for p^0 and p^1 are beta distributions with shape parmeters that sum to $n_0^0 + 2$ and $n_0^1 + 2$, respectively. Also, note the standard uniform initial prior, $U(0, 1)$, is proportional to Beta(1,1) and has an ESS of 2. Therefore, our ESS equations in (3.20) and (3.21) have the same form as the definition given by Morita et al. in (3.17) (see footnote above).

Figure 3.4 shows power priors on p^0 and p^1 using the data from Section 3.2.2, which we used to find power priors and ESS's for β_0 and β_1. The power priors on p^0 and p^1 can be closely approximated by Beta(124, 319) and Beta(132, 295) distributions, respectively (see Figure 3.4). These approximate beta priors both yield a prior ESS of 427. Recall that $n_0^0 = n_0^1 = 50$ and $a_0 = 0.85$. Then, the equations in (3.20) and (3.21) also yield ESS's of 427 analytically.

It is of interest to compare the ESS values deduced from probabilities and regression coefficients. Recall that the ESS's for β_0 and β_1 were 442 and 688, respectively. We expect the prior to contain the same amount of information for β_0 and p^0, so these values should be close, as indeed they are. When comparing ESS's for β_1 and p^1, note p^1's dependence on both β_0 and β_1 in (3.11). Therefore, we don't expect the ESS for p^1 to match either β_0 or β_1. In all of our simulations, we have found that the ESS's of the logistic regression probabilities are smaller than those of the coefficients.

It is intuitively appealing that the prior ESS's on the probabilities are directly related to their respective sample sizes. In the power prior formulation, we are given no initial expert opinion to form the priors. We use only the historical data (unless we are given an informative initial prior). So, it makes sense that only the historical data should affect the amount of information in the power prior. For example, suppose that, instead of a balanced design, we have $n_0^0 = 300$ and $n_0^1 = 700$. Then, with $a_0 = 0.85$ as before, $m^0 = 257$ and $m^1 = 597$. Intuitively the amount of information in p^0 should be less because we have less historical data corresponding to p^0.

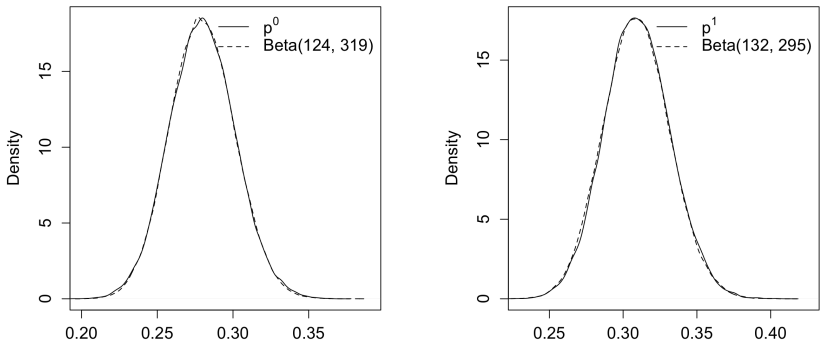


Figure 3.4: Power priors on regression probabilities found by sampling from the posterior in OpenBUGS using the same data from Section 3.2.2.

We can view a_0 choices in light of the prior ESS in a fashion similar to that in Section 3.2.2. We can do so in terms of success probabilities with minimal computational effort because the ESS of priors on the probabilities is a linear function of a_0. We can also use the graph in Figure 3.5, to choose an a_0 to match an expert-specified ESS value, again being warry of the resulting bias. For example, suppose that we wish to exploit a historical data set with $n_0^0 = 300$, $n_0^1 = 700$, and $x_0 = 0$ indicates that a patient is in the standard treatment group. Further suppose that an expert experienced with studies involving the probability of success for patients on the standard treatment recommends a maximum prior ESS of approximately 150 on p^0. Per the expert's suggestion, we would consider only a_0 values less than 0.5, which corresponds to a prior ESS of 152 on p^0.

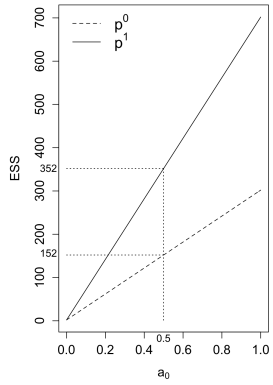


Figure 3.5: Prior ESS of success probabilities as a function of a_0.

3.2.4 Multiple Categorical Variables

The formulation in Section 3.2.3 can be extended to a categorical variable with more than two levels as well as multiple categorical predictors. For example, consider a logistic regression model for a progressive disease. Let y_0 be a binary response indicating whether or not treatment for the disease was efficacious. Let $x_0 \in \{0, 1\}$ be a dichotomous variable denoting gender and $z_0 \in \{1, 2, 3\}$ be a three-level categorical variable denoting three stages of the disease. Then, the probability that the treatment is successful for the i^{th} patient is

$$p_i = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_{0i} - \beta_2 z_{02i} - \beta_3 z_{03i})} \tag{3.22}$$

where z_{02} is 1 if the ith patient is in stage 2 of the disease and 0 otherwise, and z_{03} is 1 if the ith patient is in stage 3 of the disease and 0 otherwise.

Similar to Section 3.2.3, we partition the historical data into four subvectors, $\mathbf{y}_0^j$, with corresponding sample sizes n_0^j, $j = 1, \ldots, 4$, given the designs shown in Table 3.1. Also shown in Table 3.1 are the four probabilities p^0, p^1, p^2, p^3 corresponding to each partition. These probabilities are mutually conditionally independent (conditioned on the covariates) because the partitioned sub-vectors of $\mathbf{y}_0$ do not overlap. Similar to above, we will obtain four beta distributions from which we demonstrate the use of ESS to determine the effect of a_0.

Table 3.1: Historical data partition with corresponding probabilities.

x_0	z_{02}	z_{03}	Probabilities
0	0	0	$p^0 = [1 + \exp(-\beta_0)]^{-1}$
1	0	0	$p^1 = [1 + \exp(-\beta_0 - \beta_1)]^{-1}$
0	1	0	$p^2 = [1 + \exp(-\beta_0 - \beta_2)]^{-1}$
0	0	1	$p^3 = [1 + \exp(-\beta_0 - \beta_3)]^{-1}$

Then, similar to the formulation in (3.16), the power prior for the four distinct prob-

abilities is

$$\pi(p^0, p^1, p^2, p^3 \mid \mathbf{D}_0, a_0) = \prod_{j=0}^{3} \prod_{i=1}^{n_0^j} (p^j)^{a_0 y_{0i}^j} (1-p^j)^{a_0(1-y_{0i}^j)} \mathrm{I}_{(0,1)}(p^j). \tag{3.23}$$

This yields four independent beta distributions corresponding to the four products in (3.23):

$$p^j \mid \mathbf{y}_0^j, a_0 \sim \text{Beta}\left(a_0 \sum y_{0i}^j + 1, a_0 \sum (1 - y_{0i}^j) + 1\right), \quad j = 0, \ldots, 3. \tag{3.24}$$

Prior ESS values can be calculated for the four probabilities by summing their respective beta shape parameters. The ESS for the full power prior can be found using the formula in (3.17) given by Morita et al. [36]. As before, the prior ESS's can be used to determine the an appropriate a_0 value.

3.3 Constraining Power Parameters to Comply with FDA Guidelines

In this section, we present a method for evaluating choices for fixed a_0 in accordance with FDA guidelines on Bayesian analyses [2] and illustrate with two examples. Suppose we have study success criteria that requires the slope of the logistic regression to be greater than β^* with posterior probability at least ξ. The FDA guidance recommends that the prior probability of success be less than ξ:

> FDA recommends you evaluate the prior probability of your study claim if you are using an informative prior distribution. This is the probability of the study claim before seeing any new data, and it should not be too high. What constitutes "too high" is a case-by-case decision. In particular, we recommend the prior probability not be as high as the success criterion for the posterior probability.

It follows that we should eliminate choices of a_0 which yield prior probabilities of success in excess of the study success criteria.

3.3.1 Using FDA Criterion in Logistic Regression

To illustrate, consider a logistic regression like that in Section 3.2.2 with one dichotomous variable, x, denoting whether or not the patient received treatment. Suppose

we consider the treatment efficacious if $\beta_1 > 0$ with 95% posterior probability. The power prior on $\boldsymbol{\beta}$ is given in (3.13). Using MCMC methods, we simulate from the distribution in (3.13) and calculate the prior probability that $\beta_1 > 0$, for various values of a_0. We should rule out use of a_0 values which yield a prior probability for $\beta_1 > 0$ in excess of 95%.

We simulate a data set containing 600 points with $(\beta_0, \beta_1) = (-1, 0.4)$. We treat this as the historical data to be utilized in the current experiment. These probabilities are found using the "step" function in OpenBUGS and are plotted against various a_0 values in Figure 3.6. The dashed horizontal line represents a prior probability of 95% which corresponds to a power parameter of 0.5. Thus, power priors that take on a_0 values greater than 0.5 would violate FDA recommendations and should be avoided.

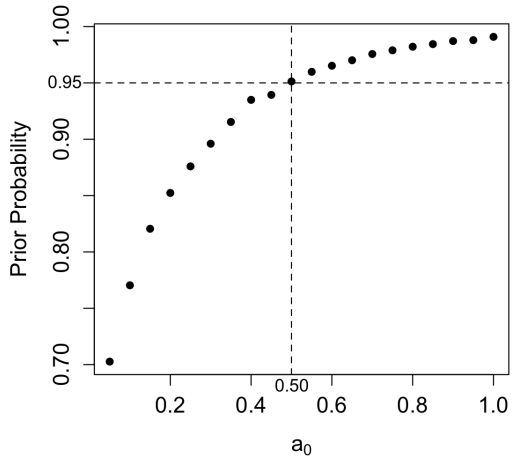


Figure 3.6: Prior probabilities of success as a function of a_0. The dotted lines represent a prior probability of 95% at $a_0 = 0.5$.

3.3.2 Using FDA Criterion with Poisson Data

As another example in determining a single a_0, we use the FDA guidance to select a fixed power parameter for Poisson data. Also, we look at the variance of the historical data under the Poisson model noting the attenuation effect of the power parameter. Suppose we have historical data, D_0, consisting of n_0 Poisson observations

$$y_{0i} \sim \text{Poisson}\,(\lambda)\,, \quad i = 1, 2, ..., n_0.$$

For the initial prior $\pi_0(\lambda)$, we assume a conjugate gamma distribution with shape parameters equal to 0.001. This has probability distribution function

$$\pi_0(\lambda) \propto \lambda^{0.001-1} \exp(-0.001\lambda) \approx \frac{1}{\lambda}. \tag{3.25}$$

Then the power prior is, for some $0 < a_0 < 1$,

$$\begin{aligned} \pi(\lambda \mid D_0, a_0) &\propto \left[\frac{\lambda^{\sum_{i=1}^{n_0} y_{0i}} \exp(-n_0\lambda)}{y_{01}! y_{02}! \times \cdots \times y_{0n_0}!} \right]^{a_0} \frac{1}{\lambda} \\ &\propto \lambda^{a_0 \sum_{i=1}^{n_0} y_{0i} - 1} \exp[-a_0 n_0 \lambda]. \end{aligned} \tag{3.26}$$

This is a gamma(α, β) distribution with $\alpha = a_0 \sum_{i=1}^{n_0} y_{0i}$ and $\beta = a_0 n_0$.

Note that the variance of the historical data is λ with maximum likelihood estimate $\hat{\lambda} = n_0^{-1} \sum_{i=1}^{n_0} y_{0i}$. The variance of λ under the power prior model is

$$\begin{aligned} \text{Var}(\lambda \mid D_0, a_0) &= \frac{\sum_{i=1}^{n_0} y_{0i}}{a_0 n_0^2} \\ &= \frac{\hat{\lambda}}{a_0 n_0}. \end{aligned}$$

Therefore, the affect of a_0 is to attenuate the historical data variance.

Similar to the logistic regression case, we can look at a_0 in light of the FDA guidelines. Recall that the FDA recommends the prior probability of success be less than the posterior success criteria. Suppose that we are planning an experiment in which we wish to show with 90% posterior certainty that the rate λ of current Poisson data is greater than 1.

We simulate 1000 historical Poisson data and use the power prior in (3.26) for our prior on λ. In order to adhere to FDA guidelines, the prior probability $P(\lambda > 1)$ must be less than 90%. The prior probability is simple to compute using the "pgamma" function in R. Figure 3.7 shows that values of a_0 greater than 0.2 yield prior probabilities $P(\lambda > 1)$ greater than 90%. Therefore, the choice for a_0 is narrowed to values less than 0.20. If the success criterion was 95%, choices for a_0 would be narrowed to a_0's less than 0.3.

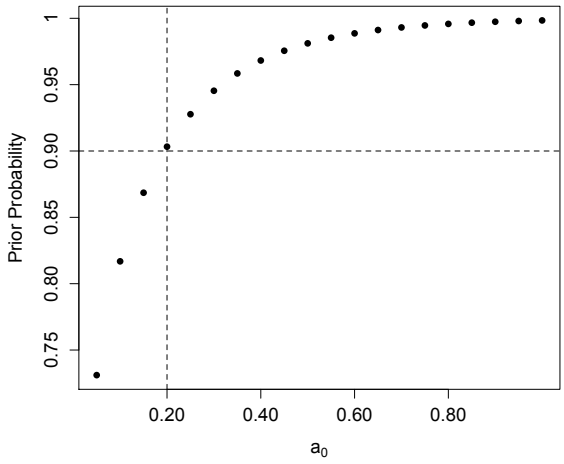


Figure 3.7: The power parameter a_0's effect on the prior probability $P(\lambda > 1)$ with dashed lines representing 90% probability and corresponding $a_0 = 0.2$.

3.3.3 Prior Predictive Probability under the FDA Guidance

In this section, we consider a prior success probability that is a function of the data, necessitating use of the prior predictive distribution. Suppose we are planning an experiment with n patients who will be receiving an experimental treatment. Assume that the planned experiment will be considered successful if less than some fraction, δ, of patients on the experimental treatment experience an adverse effect within 5 years with 90% posterior certainty. Here, δ is chosen to indicate a smaller percentage than that for some standard treatment. Let y be the number of patients out of n who experience an adverse effect in the 5-year study period and $y^* = \delta n$. Thus, interest is in the probability that $y \leq \delta n$. To comply with FDA regulations, the probability that y is less than y^* should not exceed 90% *a priori*. Unlike the previous two examples, here we require probability calculation with respect to the prior predictive distribution because the probability is on a function of the unobserved data, y.

Suppose there is a historical data set of size n_0, yielding a count of y_0 patients on a similar experimental treatment who experienced adverse effects in a 5-year duration. Denote this historical data $D_0 = (y_0, n_0)$. Here, $y_0 \sim \text{Bin}(n_0, \theta)$. Then, to utilize the

historical data, we use a power prior with an Beta(α, β) initial prior, chosen, here, for convenience (conjugacy). Thus, we have the power prior

$$\pi(\theta \mid D_0, a_0) \propto \theta^{a_0 \sum_{i=1}^{n_0} y_{0i} + \alpha - 1} (1-\theta)^{a_0 n_0 - a_0 \sum_{i=1}^{n_0} y_{0i} + \beta - 1}. \tag{3.27}$$

This is a Beta(ξ, η) distribution where $\xi = a_0 \sum_{i=1}^{n_0} y_{0i} + \alpha$ and $\eta = a_0 n_0 - a_0 \sum_{i=1}^{n_0} y_{0i} + \beta$. Thus, we have the prior predictive distribution

$$\begin{aligned} p(y) &= \frac{\pi(y \mid \theta)\pi(\theta \mid D_0, a_0)}{\pi(\theta \mid y)} \\ &\propto \binom{n}{y} \frac{\mathrm{B}(\xi, \eta)}{\mathrm{B}(y + \xi, n - y + \eta)} \end{aligned} \tag{3.28}$$

where B$(\cdot, \cdot)$ is the Beta function. Therefore, (3.28) is a Beta-Binomial distribution.

We find the prior predictive density in OpenBUGS by specifying the beta power prior in (3.27) with the current likelihood. However, we do not give OpenBUGS any data for the current likelihood so that the power prior drives all information on y, yielding the prior predictive distribution for y. From there, we find the prior predictive probability that $y < y^*$ using the "step" function in OpenBUGS.

To illlustrate, suppose we are planning a current study to compare the prevalence of an adverse reaction in an experimental treatment to a standard therapy. We wish to show that the number of patients out of $n = 300$ on the experimental treatment who experience an adverse reaction, y, is less than 10% of n with 90% certainty a posteriori. We simulate y_0 from a Bin$(500, 0.072)$. This means that the power prior (3.27) reflects information in favor of our current success criterion because the expected value of y_0 is 36, which is indeed less than 10% of 500. Therefore, it is likely that some high values of a_0 may yield prior probabilities exceeding 90%. This is because larger a_0 would incorporate more evidence that the number of adverse reactions is less than 10% of n_0. We let the hyper-parameters be $\alpha = \beta = 1$. Figure 3.8 shows the prior predictive probability that y is greater than 30 as a function of a_0. Values of a_0 greater than 0.3 yield success probabilities greater than 90% and violate FDA rdsegulations.

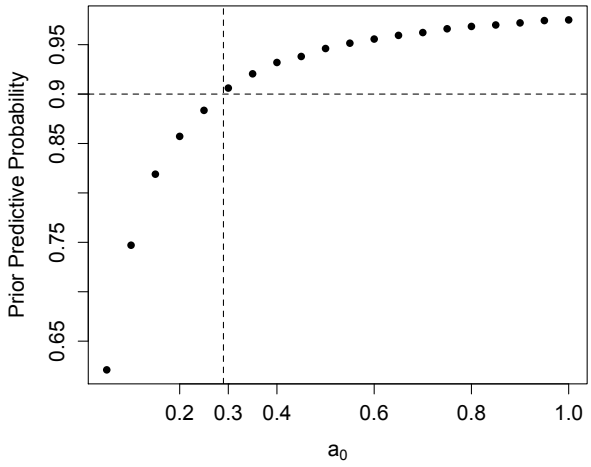


Figure 3.8: Prior predictive probability that y is less than 5 for various a_0's. A dotted line is drawn at 90% probability and at the corresponding a_0 value of appoximately 0.3.

3.4 Power Parameter Choices in Multiple Studies

The power prior formulation can be generalized to include more than one historical study. In this section, we consider power parameter for such models, which are referred to as multiplie-studies power priors. Here, a vector of power parameters, $\mathbf{a}_0 = (a_1, ..., a_K)$ for K studies is of interest. In Section 3.4.2, we propose a method to constrain power parameter choices based on the prior ESS, FDA guidelines, and expert information. We present an example using logistic regression data in Section 3.4.3.

Reitbergen et al. [45] propose eliciting a ranking of the historical studies according to their commensurability with the current data. In their discussion, the commensurability between the historical and current trials is based on study characteristics chosen by the expert as most important.

With this in mind, after ranking the historical trials through expert opinion, we propose rejecting the power-parameter vectors which violate FDA guidelines and/or prior ESS constraints. With the remaining power parameters, we again incorporate expert information through the prior ESS.

3.4.1 Multiple Studies Power Prior

In this section, we briefly introduce the form and notation for multiple studies power priors and conclude with a graphical illustration. Suppose there are K_0 historical studies, with kth study summary

$$D_{0k} = (n_{0k}, \mathbf{y}_{0k}, \mathbf{X}_{0k}), k = 1, \ldots, K_0.$$

Let $\mathcal{D}_0 = (D_{01}, \ldots, D_{0K_0})$. We can have multiple weight parameters, $\mathbf{a}_0 = (a_{01}, \ldots, a_{0K_0})$ where a_{0k} is the weight for the kth study. Then the power prior incorporating the K_0 studies is defined

$$\pi(\boldsymbol{\theta}|\mathcal{D}_0, a_0) \propto \left(\prod_{k=1}^{K_0} [l(\boldsymbol{\theta}|D_{0k})]^{a_{0k}} \right) \pi_0(\boldsymbol{\theta}). \tag{3.29}$$

As an example of the power prior formulation with multiple historical studies, we consider logistic regression with three historical studies, where x_0 is a binary predictor. We give β_0 and β_1 independent normal initial priors, denoted $\phi_0(\mu, \sigma^2)$. Then the power prior is

$$\begin{aligned} \pi(\boldsymbol{\beta} \mid \mathcal{D}_0, \mathbf{a}_0) = \prod_{k=1}^{3} & \left(\frac{1}{1 + \exp(-\beta_0 - \beta_1 x_{0k})} \right)^{a_{0k} \sum_{i=1}^{n_{0k}} y_{0ki}} \\ & \times \left(1 - \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_{0k})} \right)^{a_{0k} \sum_{i=1}^{n_{0k}} (1 - y_{0ki})} \\ & \times \phi_0(\mu, \sigma^2) \phi_0(\mu, \sigma^2). \end{aligned} \tag{3.30}$$

Here, we take $\mu = 0$ and $\sigma = 10$. We simulate data from 3 historical logistic regression models with $\beta_0 = -1$ and $\beta_1 = 0.16$. In each data set, half of the data corresponds to $x_0 = 0$ and half to $x_0 = 1$. We generate 3000 points from the first historical model, 2000 from the second, and 2500 from the third. We use the power parameters $\mathbf{a}_0 = (0.6, 0.65, 0.75)$. Figure 3.9 shows the power priors on β_0 and β_1 incorporating the three historical studies.

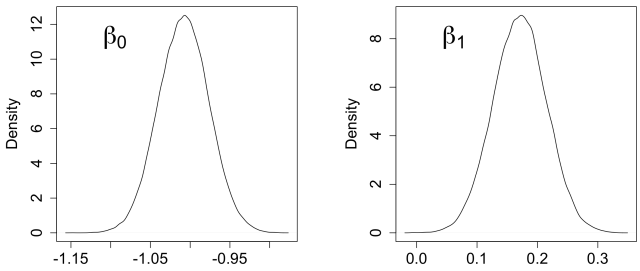


Figure 3.9: Power priors on β_0 and β_1 from multiple studies with $\mathbf{a}_0 = (0.6, 0.65, 0.75)$.

3.4.2 Method for Multiple Power Parameters Selection

Reitbergen et al. [45] propose eliciting a ranking of the historical studies according to their commensurability with the current data. With this in mind, after ranking the historical trials through expert opinion, we propose rejecting the power-parameter vectors which violate FDA guidelines and/or prior ESS constraints. With the remaining power parameters, we again incorporate expert information through the prior ESS.

To arrive at a prior ESS value, we elicit a hypothetical number of data points from the historical trials that the expert would add to the current trial, based on their commensurability. In a clinical trial setting, such a question may be "How many patients from the historical trial would you incorporate into this trial based on the overall commensurability?" This commensurability might include but should not be limited to similarities between the patient populations. Being wary of potential bias, we treat this number as the expert's maximum ESS for the power prior. Note that, while it would be ideal to match an elicited ESS for each historical study to a corresponding power parameter, it is not clear that we can split the multiple studies power prior on joint parameters in this way.[5] Therefore, our elicited prior ESS should be one value, which can be obtained by summing study-specific ESS's if desired.

Suppose an expert ranks the studies $1, 2, \ldots, K$, in increasing order of commensurability. We assume there are no ties. Let a_{0j} be the power parameter used for the study with

[5] Note that in Section 3.4.4, we do "split" the power prior and elicit trial-specifc ESS's. This is possible, in part, because the power prior is on only one parameter.

rank j, $j = 1, \ldots, K$. We assume $a_{01} < a_{02} < \ldots < a_{0K}$. Let $\mathbf{a}_0 \equiv (a_{01}, a_{02}, \ldots a_{0K})$. While the power parameter can be any real number in theory, for efficiency, suppose $a_{0i} \in (0, 0.05, 0.1, 0.15, ..., 1)$ for $i = 1, \ldots, K$. With this assumption, we simulate all possible $\mathbf{a}_0$'s such that $a_{01} < a_{02} < \cdots < a_{0K}$.

We find the prior probability of success corresponding to each vector $\mathbf{a}_0$. Some values of the $\mathbf{a}_0$'s can be eliminated by use of the previously mentioned FDA guidelines. We illustrate this approach with an example in the next section.

3.4.3 Multiple Power Parameter Selection in Logistic Regression

In this example, we use the FDA guidelines, the prior ESS, and expert information for selection of power parameters in a logistic regression model with multiple historical studies. We simulate 350 data points from each of three logistic regression models with one binary predictor:

$$y_{0ik} \sim \text{Bernoulli}(p_{0ik}) \tag{3.31}$$

where

$$\text{logit}(p_{0ik}) = \beta_0 + \beta_1 x_{0ik}, \quad k = 1, 2, 3$$

and $(\beta_0, \beta_1) = (-1, 0.25)$. These are treated as the historical models. We construct all models to be balanced, so that, in each of the trials, half of the data corresponds to a zero covariate value. The power prior for this problem is shown in (3.30). Here, we take μ to be normal with mean 0 and variance 1000, and σ to be uniform over 0 to 10. Suppose that we are planning a current experiment with 500 data points for which we wish to utilize the historical data in (3.31). For simplicity, suppose that the experiment is successful if $\beta_1 > 0$ with 95% posterior certainty.

Assume an expert has ordered the historical studies and we have corresponding power parameters $a_{01} < a_{02} < a_{03}$. From the expert's ranking, we form a grid of power

parameter vectors such that:

$$\begin{aligned} a_{01} &\in (0, 0.05, 0.1, 0.15, \ldots, 1), \\ a_{02} &\in [a_{01}, a_{01} + 0.05, a_{01} + 0.1, a_{01} + 0.15, \ldots, 1), \end{aligned} \tag{3.32}$$

and

$$a_{03} \in [a_{02}, a_{02} + 0.05, a_{02} + 0.1, a_{02} + 0.15, \ldots, 1).$$

This yields a grid consisting of 969 vectors of power parameters. As discussed in Section 3.3, the FDA recommends that the prior probability of success not exceed the posterior success criteria. We use the "step" function in OpenBUGS along with the functional forms of the power priors to find the corresponding prior probabilities of success. Out of the 969 $\mathbf{a}_0$-vectors, 748 yield prior success probabilities in excess of 95%. Therefore, we eliminate 77% of the vectors, leaving 221 in consideration.

We now utilize the prior ESS, as discussed in Section 3.2, to futher constrain choices for $\mathbf{a}_0$. However, here, note that our historical models were simulated such that $\beta_1 > 0$. In this way, the historical data provides evidence in favor of the current success criterion. Therefore, $\mathbf{a}_0$ choices that result in large contributions of historical information yield large prior success probabilities. Thus, as we will see, for this example all $\mathbf{a}_0$'s yielding extreme prior ESS's have already been eliminated based on the FDA constraint. Furthermore, we proclude use of any $\mathbf{a}_0$ yielding prior ESS values for (β_0, β_1) larger than the current study's sample size, the corresponding $\mathbf{a}_0$'s would be automatically disqualified.

We find the ESS's for the power priors corresponding to the remaining 221 $\mathbf{a}_0$'s using the R program `ESS_RegressionCalculator` detailed in Section 3.2.2. Figure 3.10 shows box plots of the prior ESS's. Notice, in particular, a power parameter vector of $\mathbf{a}_0 = (0.2, 0.35, 0.45)$ yields a power prior with an ESS of 129.73 for the power prior with (β_0, β_1). This is the maximum ESS that the power prior can have while complying with FDA regulations.

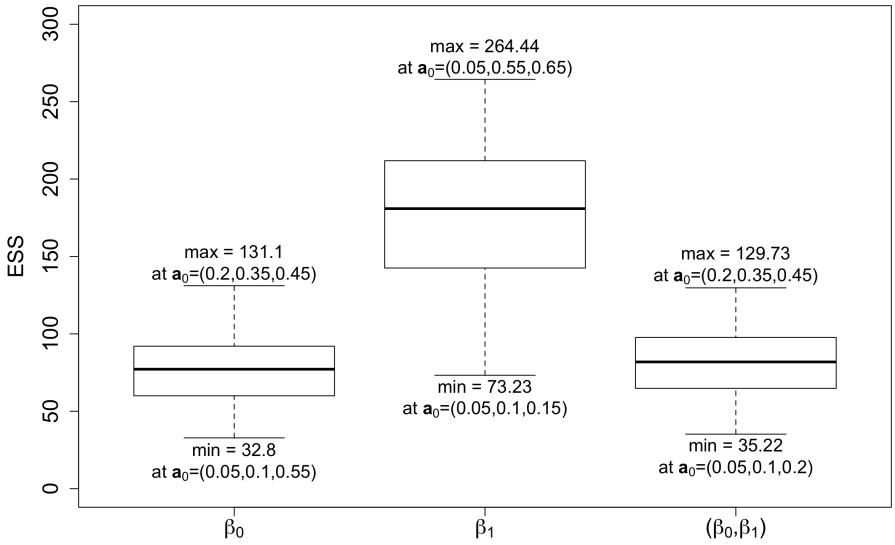


Figure 3.10: Box plots of the prior ESS values for $\boldsymbol{\beta}$, β_0, and β_1 resulting from multiple-studies power priors with varying $\mathbf{a}_0$-vectors.

Suppose it is decided to limit the effect of historical information in the model by requiring that the ESS be no more than, say, 1/5 of the planned study size. We find the $\mathbf{a}_0$'s corresponding to power priors with ESS's within a ± 5 margin of 100. This narrows the choice for $\mathbf{a}_0$ down to 15 vectors, about which we suggest further consulting the expert. The resulting vectors are shown in Table 3.2 below. The expert may believe that Trial 1 has little commensurability with the current study, while Trial 3 is very relevant. In this case, the last set of power parameters in Table 3.2 would be appropriate. The marginal prior ESS's on β_0 and β_1 and the prior probabilities of study success are also provided.

Note that the prior probability of study success for all options in Table 3.2 are close to 95%. This is the result of two choices. First, the historical data was simulated in favor of the study success criterion. Increasing the effect of the historical likelihood increases the prior probability of study success. Second, the expert's ESS was close to the maximum ESS cut-off before violating FDA guidelines. Note that, after viewing the table, the expert may wish to modify her opinion. The statistician can also increase the margin around the elicited upper bound for the prior ESS to include more options for $\mathbf{a}_0$.

Table 3.2: Power parameters with corresponding prior ESS's for $\boldsymbol{\beta}$ within a ± 5 margin of the expert-elicited upper bound on the prior ESS.

a_{01}	a_{02}	a_{03}	ESS($\boldsymbol{\beta}$)	ESS(β_0)	ESS(β_1)	$P(\beta_1 > 0 \mid D_0)$
0.20	0.25	0.35	107	108	211	0.935
0.20	0.25	0.40	108	107	220	0.944
0.15	0.30	0.45	107	104	225	0.949
0.15	0.30	0.50	108	104	231	0.946
0.15	0.35	0.50	110	107	228	0.940
0.10	0.40	0.50	109	104	231	0.941
0.15	0.30	0.55	111	106	234	0.949
0.10	0.40	0.55	110	105	233	0.941
0.05	0.50	0.55	110	105	235	0.938
0.15	0.30	0.60	110	104	236	0.950
0.05	0.50	0.60	113	106	247	0.941
0.10	0.40	0.65	109	101	239	0.944
0.05	0.50	0.65	110	103	243	0.944
0.05	0.50	0.70	112	104	249	0.940
0.05	0.45	0.90	107	97	246	0.948

3.4.4 Multiple Power Parameters with Binomial Data

In this section, we consider a non-regression binomial model with a beta mulitple-studies power prior. Specifically, suppose we have K historical data sets that all study the success probability, θ for some treatment. Assume they have sample sizes $n_{01}, \ldots, n_{0K}$ with $k_{01}, \ldots, k_{0K}$ successes, respectively. Then, using a U(0,1) initial prior, the power prior on θ that incorporates the K studies is

$$\begin{aligned}
\pi\left(\theta \mid \mathbf{D}_0, \mathbf{a}_0\right) &\propto \theta^{a_{01}k_{01}}\left(1-\theta\right)^{a_{01}(n_{01}-k_{01})} \\
&\quad \times \theta^{a_{02}k_{02}}\left(1-\theta\right)^{a_{02}(n_{02}-k_{02})} \\
&\quad \vdots \\
&\quad \times \theta^{a_{0K}k_{0K}}\left(1-\theta\right)^{a_{0K}(n_{0K}-k_{0K})} \mathbf{I}_{[0,1]}(\theta) \\
&= \theta^{\sum_{i=1}^{K} a_{0i}k_{0i}}\left(1-\theta\right)^{\sum_{i=1}^{K} a_{0i}(n_{0i}-k_{0i})} \mathbf{I}_{[0,1]}(\theta).
\end{aligned} \tag{3.33}$$

This is a beta prior with ESS equal to

$$m = a_{01}n_{01} + \cdots + a_{0K}n_{0K} + 2. \tag{3.34}$$

By dividing the overall ESS in (3.34) by the number of studies, K, we propose that a prior ESS value can be elicited from an expert for each individual trial and treated as

$$m_k = a_{0k}n_{0k} + \frac{2}{K}, \quad k = 1, \ldots, K. \tag{3.35}$$

This is intuitively appealing. Each historical study contributes a fraction of its sample size to the total ESS, where the fraction in determined by the value of a_0. The addition of $2/K$ is minimal, especially for large K and large sample sizes, and represents the prior ESS of the diffuse initial prior.

We propose the following method to elicit trial-specific power parameters in beta/binomial models. Set m to be the maximum overall prior ESS chosen by the statistician. We recommend m be at most equal to the planned sample size for the current experiment. The maximum value of m can be better determined by following the formulation in Section 3.4.3. In particular, recall that after eliminating $\mathbf{a}_0$'s that yield prior probabilities in violation of FDA guidelines, the remaining $\mathbf{a}_0$'s yield a maximum overall prior ESS. We propose treating m as that maximum ESS corresponding to the remaining $\mathbf{a}_0$ vectors. Letting the expert know m, we elicit individual maximum ESS's $m_1, \ldots, m_K$ as a rough measure of commensurability for each historical study so that the sum is less than m. We treat these m_i's as upper bounds on trial-specific prior ESS's, and repeat a process similar to the method in Section 3.4.2.

To illustrate, consider three historical binomial data sets of sample sizes, n_{01}, n_{02}, n_{03}, of 600, 700, and 750, respectively. Suppose we are planning a similar current study with $n = 700$ for which we wish to incorporate the historical data through a multiple-studies power prior. Suppose that, using the FDA criteria, we have eliminated possible $\mathbf{a}_0$ vectors and find that m should be 500 at a maximum.

Based on the commensurability between each historical trial with the current trial, suppose the expert gives trial-specific maximum ESS's of 300, 75, and 100 for m_1, m_2, and m_3, respectively. This implies that the corresponding power parameters can be found through (3.35) via

$$\begin{aligned} 300 &= 600a_{01} + 2/3, \\ 75 &= 700a_{02} + 2/3, \end{aligned} \tag{3.36}$$

and

$$100 = 750a_{03} + 2/3.$$

This yields a power parameter vector of $\mathbf{a}_0 = (0.5, 0.11, 0.13)$.

3.5 Power Parameter Selection with Multiple Experts

In this section, we assume we have multiple experts from which we elicit information regarding the vector of power parameters in power priors with multiple studies. We present an example from Reitburgen et al. [44] and propose a technique to further narrow $\mathbf{a}_0$ choices. In their 2014 paper, Reitbergen et al. state that "...the specification of study weights depends heavily on study specific substantial aspects, rather than statistical ones, making (clinical) expert knowledge crucial in this process." Mathematical and/or behavioural aggregation methods can be employed to reconcile on a set of ranks between the experts. O'Hagan et al. [41] have a nice chapter on these methods, which can be split into two main categories: mathematical and behavioral aggregation. Behavioral methods rely on techniques to elicit information from experts as a group. Mathematical methods do not require a group consensus on the information, but involve mathematically combining expert opinions. In the example below, Reitburgen et al. use the well-known Delphi method, a behavioral approach.

Rietbergen et al. [44] provide a detailed example of eliciting $\mathbf{a}_0$ from multiple experts for a set of studies. They use the Delphi technique to reach an agreement between the experts on an $\mathbf{a}_0$ vector. A panel of four experts is consulted to judge the degree of overlap

between four historical studies and the current study, which evaluate an insulin sensitizer called Rosiglitazon. Details about their choice of experts and historical study inclusion can be found in their paper.

In [44], in the first Delphi round, the experts were asked to rank the four studies from most to least relevant. All of the experts reached consensus on the ranking in the third round. They were also asked to assign study weights along with each ranking. For example, the authors explain that a weight of 50 implies that the expert is willing to incorporate 50% of the historical data in the prior distribution for the new data. While the variability between the experts' weights decreased over the three rounds, agreement was not reached. From the written expert explanations, Rietbergen et al. found that similarities in the endpoints used for the outcome measure most motivated higher ranks and weights. The experts' written comments are provided in Rietbergen et al. [44].

When consensus is not reached, our methods can be used to narrow the power parameter vectors. We eliminate expert's vectors which automatically place the experiment in violation with FDA guidelines. We can then use the prior ESS corresponding to each $\mathbf{a}_0$'s to discuss choices with the experts.

Suppose the experts reach an agreement on $\mathbf{a}_0$. The choice of this vector may yield prior probabilities which violate the FDA recommendations or contain too much information in terms of prior ESS. To this end, we propose the following method. Suppose we have K_0 studies which correspond to K_0 power parameters

$$\mathbf{a}_0 = \{a_{01}, a_{02}, \ldots, a_{0K_0}\}.$$

Further suppose the experts have ordered the power parameters from most relevant to least relevant have agreed upon fixed values.

When the agreed-upon fixed $\mathbf{a}_0$ vector leads to violation of FDA guidelines or excessive prior ESS, we recommend eliminating a_0 values beginning with smallest. In effect, we have K_0 choices of $\mathbf{a}_0$ vectors. As in the example in Section 3.1.2, we can plot the

prior ESS and/or prior probability against the K_0 power parameter vectors. To illustrate, consider the logistic regression with multiple historical studies in example in Figure 3.9. Suppose that experts have selected the vector

$$\mathbf{a}_0 = \{0.75, 0.65, 0.6\}. \tag{3.37}$$

Further, suppose that a successful trial will yield a $\Pr(\beta_1 > 0)$ with 95% certainty *a posteriori*. The FDA recommends the prior $\Pr(\beta_1 > 1)$ not exceed 95%. The power parameter vector in (3.37) recommended by the experts yields a prior $\Pr(\beta_1 > 1)$ of 0.996. To eliminate the smallest a_{0i}, we set $a_{03} = 0$ so that Trial 3 is excluded. This yields a power prior with prior study success probability of 0.9663. Lastly, we consider just one $a_0 = 0.75$. This power prior incorporates only Trial 1 and has prior $\Pr(\beta_1 > 0) = 0.814$.

This method is appealing because it is not computationally expensive or complex. However, this is at the cost of potentially ignoring useful information from eliminated trials.

3.6 Discussion

Power priors facilitate utilization of historical information for inference from current experiments. Power priors are typically simple in their utilization and interpretation. However, care must be taken in consideration of the power parameter, a_0. It is not recommended to give a_0 a prior distribution. Therefore, determination of a fixed a_0 requires an operational assessment of implications of such choices. We proposed three main methods for doing so by utilizing the prior ESS, the FDA guidelines on Bayesian analyses, and expert information.

In most cases, our proposed methods ruled out a large proportion of power parameter choices. In some cases, we were able to select a value of a_0 based on the particular criteria. In Section 3.4.4, we were able to "split" the $\mathbf{a}_0$ vector corresponding to the multiple-studies power prior into individual trial-specific power parameters. This allowed for elicitation of

trial-specific prior ESS's which directly implied values for each historical study's individual power parameter. Further work could include a determination of whether or not this type of "splitting" is feasible in multiple-studies power priors on joint parameters, such as that in Section 3.4.3.

CHAPTER FOUR

Sensitivity to Prior Misspecifications in the Mode-Percentile Method of Elicitation

4.1 Introduction

Prior elicitation refers to the process of converting expert opinion into a probability distribution for use in a Bayesian data analysis. One question that arises in this process is the sensitivity of the resulting priors to slight deviations in the expert's specifications, which may arise from his inability to precisely quantify his beliefs or miscommunication between the expert and statistician. See, for example, O'Hagan et al. [41], Ch. 3. In this chapter we investigate the beta-binomial model's sensitivity to imprecisely specified prior summaries.

While the goal of prior elicitation is to represent the expert's beliefs, doing so can be problematic. In their third chapter, O'Hagan et al. [41] provide a detailed overview of difficulties in eliciting expert judgements. The literature on this topic is vast, including [11], [23], [31], and [41]. Difficulties range from an expert's misunderstanding of statistical terminology to bias introduced through the method of questioning. The hope is that the inevitable misspecifications in the elicitation do not produce substantially different prior distributions used in the corresponding analysis. However, this is not always the case. In particular, imprecision in expert summaries can lead to large deviations from their beliefs.

The progression of the current chapter is as follows. In the first section, we provide a background and conceptual framework in which to consider the problem. We quantify and present graphics to visualize the extent of the misspecification in Section 4.2. We extend the development in Section 4.3 to generalize the graphs across a broad range of misspecifications. We call these graphs sensitivity indicatrices and describe their usefulness in the elicitation process. We show that the use of lower percentiles results in a more robust elicitation scheme. We conclude the chapter in Section 4.4.

4.1.1 Background

Suppose we wish to elicit a joint prior distribution for a k-dimensional parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^k$. We elicit information from an expert in order to form a prior on $\boldsymbol{\theta}$. Doing so requires the specification of a vector $\mathbf{d} \in \mathbf{D} \subseteq \mathbb{R}^q$, comprised of summaries such as a mode and various percentiles. We call $\mathbf{D}$ the *elicitation space* containing all possible elicited values.

An elicitation exercise yields a value $\mathbf{d}_E \in \mathbf{D}$. We say $\mathbf{d}_T \in \mathbf{D}$ is the expert's *true belief* if there is no refinement of the exercise, such as rephrasing questions, that would yield a different value of $\mathbf{d}_T$. We assume that $\mathbf{d}_T \in \mathbf{D}$ exists and that there is an elicitation exercise that will yield $\mathbf{d}_T$. If $\mathbf{d}_E \neq \mathbf{d}_T$, then $\mathbf{d}_E$ is misspecified.

Now, let $\mathcal{P}$ be a class of priors so that $\pi_{\mathbf{d}}(\boldsymbol{\theta}) \in \mathcal{P}$ denotes a prior for $\boldsymbol{\theta}$ determined by $\mathbf{d} \in \mathbf{D}$.[1] A fundamental question about any elicitation exercise is how does $\pi_{\mathbf{d}}(\boldsymbol{\theta})$ vary with $\mathbf{d} \in \mathbf{D}$? Clearly, a misspecified $\mathbf{d}_E$ will result in a different prior than that from $\mathbf{d}_T$. We let $\pi_{\mathbf{d}_e}(\boldsymbol{\theta}) \in \mathcal{P}$ and $\pi_{\mathbf{d}_T}(\boldsymbol{\theta}) \in \mathcal{P}$ be the priors resulting from the elicited beliefs $\mathbf{d}_E$ and the expert's true beliefs $\mathbf{d}_T$ respectively.

The prior effective sample size (ESS) is commonly used as a measure of informativeness [36] [37]. The ESS is the number of observations that would need to be added to the model in order to yield a comparable analysis using a relatively non-informative prior. For more details, see the discussion in Section 3.2 of Chapter three. In this chapter, we use the prior ESS to gauge a model's sensitivity to misspecifications by comparing the ESS of $\pi_{\mathbf{d}_T}(\boldsymbol{\theta})$ to that of a misspecified $\pi_{\mathbf{d}_E}(\boldsymbol{\theta})$ for many different values of $\mathbf{d}_T$ and a quantified level of misspecification.

[1] At times, $\mathbf{d}_E$ will not yield a unique prior. In these cases, we use the prior with the smallest ESS. (See Section 3.2 of Chapter three for details on prior ESS.)

4.1.2 Problem

In this chapter, we consider the problem of prior misspecification in the mode-percentile elicitation method for a Bernoulli data model with a beta prior. We let θ be the Bernoulli success probability and take $\theta \sim \text{Beta}(\alpha, \beta)$.

In the mode-percentile method, we elicit a mode m and upper percentile p from an expert. An expert might specify that θ is most likely m and less than p. In turn, we treat m and p as the mode and upper percentile of a Beta(α, β) prior. Note that the map from the mode-percentile space into the standard parameter space is not injective: for some mode-percentile specifications, more than one (α, β) combination exists with the given mode and percentile. Further, it is known that the map is one-to-two in these places. The map can be made well-defined by always taking the value of the map to be the combination (α, β) whose sum, the ESS, is smaller. We assume this here.

Therefore, for a well-defined region of the elicitation space, we have a well-defined injective map $f : (0,1)^2 \longrightarrow (1,\infty)^2 = \boldsymbol{\Theta}$ such that $(m,p) \overset{f}{\longmapsto} (\alpha,\beta)$. However, the map f does not have an explicit representation and therefore, for a given m and p, we numerically solve for the corresponding α and β according to the following scheme. For a mode m and percentile p, we choose ξ such that $\Pr(\theta \leq p) = \xi$, say 0.85. We set

$$\xi = \int_0^p \pi_{\mathbf{d}}(\theta \mid \alpha, \beta) d\theta \tag{4.1}$$

where $\pi_{\mathbf{d}}$ is the probability density for the beta distribution. We set m equal to the mode of the distribution:

$$m = \frac{\alpha - 1}{\alpha + \beta - 2}. \tag{4.2}$$

This restricts the standard parameter space so that $\alpha, \beta > 1$, a minor restriction to ensure unimodality. We then solve for α in (4.2) and substitute this solution in (4.1). We numerically solve for β by finding the root of the function

$$g(\beta) = \xi - \int_0^p \pi_{\mathbf{d}}\left(\theta \middle| \frac{1 + m(\beta - 2)}{1 - m}, \beta\right) d\theta. \tag{4.3}$$

Note that this is a simple process since the integral is the cumulative distribution function of the beta distribution, for which there is no closed-form, but there are highly efficient and accurate implementations. We substitute the solution for β into (4.2) to find α.

We denote the true beliefs (m_T, p_T) and corresponding shape parameters (α_T, β_T). If an expert's (m_E, p_E) does not equal (m_T, p_T), then (m_E, p_E) is misspecified. For example, suppose that an expert has true beliefs $(m_T, p_T) = (0.42, 0.51)$ but specifies $(m_E, p_E) = (0.4, 0.5)$. Will the misspecification really make a difference in the model? In the subsequent development, we supress the subscript E's, denoting elicited values, for ease of notation.

4.2 Visualizing Misspecification

We begin by considering a true belief (m_T, p_T) in the elicitation space. Figure 4.1 illustrates the map f (see (4.1) and (4.2)). We indicate (m_T, p_T) as a asterick, which is mapped via f to a resulting (α, β) in standard parameter space. A misspecification can occur in some range around (m_T, p_T). In order to systematically investigate the effect of misspecification, we need to have a way to quantify and limit the extent of misspecification. A natural way to do this is to place a disk (of uniform radius) in the elicitation space so that every point in the disk represents a potential misspecified value. In Figure 4.1, we form a circle around (m_T, p_T) with radius $r = 0.02$. Every point in the circle is a misspecification of (m_T, p_T). The graphic immediately suggests the questions: "what does the image of the disk look like under f, and how does the ESS vary over this deformed region?"

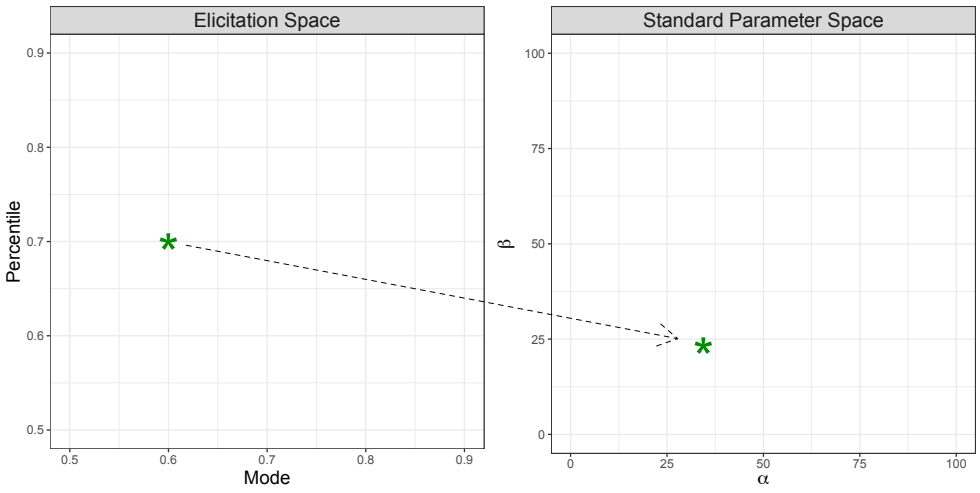


Figure 4.1: A true belief in elicitation space mapped to resulting shape parameters.

Answering these questions mathematically is not possible without an explicit form of f. Therefore, we solve the problem computationally as follows. We take a mesh of points along the boundary of the disk and form a path between them. We take each point in the mesh and follow the same map f to the standard parameter space, tying each point together in the same order as in the original space. Figure 4.2 shows the mesh in the elicitation space mapped to the standard parameter space.

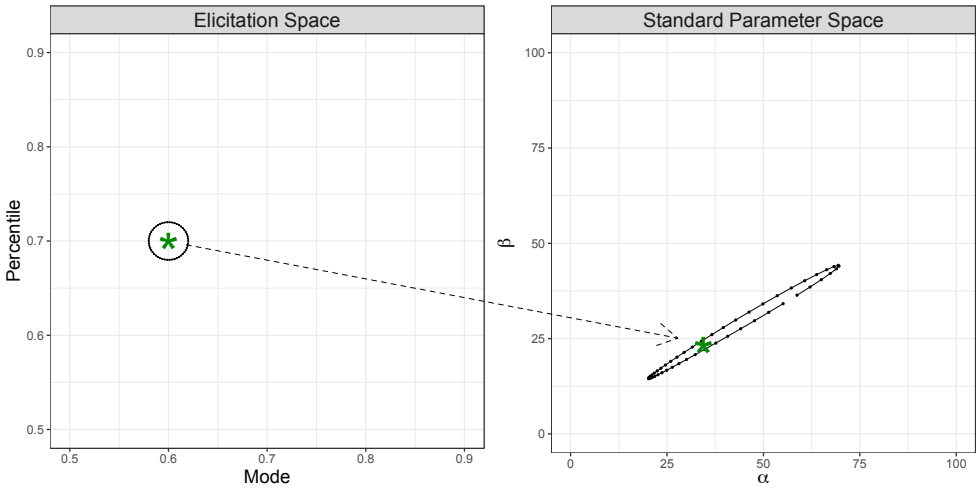


Figure 4.2: A wire frame of points around true beliefs mapped to the standard parameter space.

The disk and its resulting image $\mathcal{I}$ are shown in Figure 4.4. The image convincingly suggests that the image forms a convex set in the standard parameter space and the image is elliptical, and we operate under this assumption for the rest of this chapter.

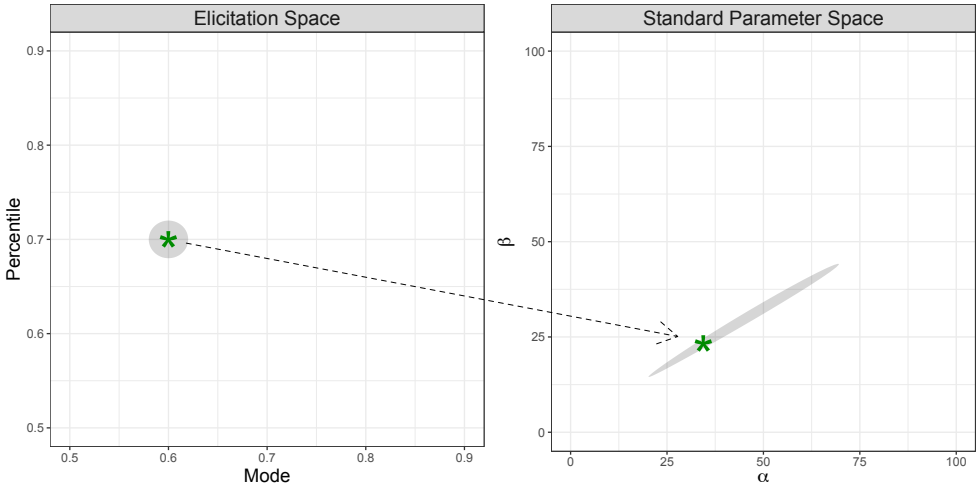


Figure 4.3: A circle representing misspecifications of (m_T, p_T) mapped to its image $\mathcal{I}$.

4.2.1 Quantifying Misspecification

We begin this section with a brief review of the necessary algebraic structures. A norm on a real vector space $\mathcal{V}$ is a function $||\cdot|| : \mathcal{V} \to \mathbb{R}$ satisfying the following three properties for any $a \in R$ and $\mathbf{v}, \mathbf{u} \in \mathcal{V}$:

(1) if $||\mathbf{v}|| = 0$, then $\mathbf{v} = \mathbf{0}$, the additive identity in $\mathcal{V}$,

(2) $||a\mathbf{v}|| = |a|||\mathbf{v}||$, and

(3) $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$ (the triangle inequality).

Norms provide a notion of length for vectors. A similarly related concept, that of a metric, provides a notion of distance between them. A metric δ on a real vector space $\mathcal{V}$ is a function $\delta : \mathcal{V} \times \mathcal{V} \to R$ satisfying the following properties for any vectors $\mathbf{v}, \mathbf{u}, \mathbf{w} \in \mathcal{V}$:

(1) $\delta(\mathbf{v}, \mathbf{u}) = 0$ if and only if $\mathbf{v} = \mathbf{u}$,

(2) $\delta(\mathbf{v}, \mathbf{u}) = \delta(\mathbf{u}, \mathbf{v})$, and

(3) $\delta(\mathbf{v}, \mathbf{w}) \leq \delta(\mathbf{v}, \mathbf{u}) + \delta(\mathbf{u}, \mathbf{w})$.

Norms naturally induce metrics in the following way: if $||\cdot||$ is a norm on $\mathcal{V}$, then $\delta(\mathbf{v}, \mathbf{u}) = ||\mathbf{v} - \mathbf{u}||$ is a metric on $\mathcal{V}$. A ball in $\mathcal{V}$ of radius $r > 0$ centered at a vector $\mathbf{v}$ is the collection of all points in $\mathcal{V}$ at most r units from $\mathbf{v}$, namely $\mathcal{B}_{\mathbf{v},r} = \{\mathbf{u} \in \mathcal{V} : \delta(\mathbf{v}, \mathbf{u}) \leq r\}$.

Up until this point, our language has suggested the use of the Euclidean norm: $||\mathbf{v}||_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$. We wish to describe the extent of the misspecification, but it is difficult to the describe distances using the Euclidean norm. Indeed, in an ℓ_2 ball, like that in Figure 4.2 with radius r, for every point (m, p),

$$\sqrt{(m - m_T)^2 + (p - p_T)^2} \leq r. \tag{4.4}$$

To better describe the extent of the misspecification, we utilize the ℓ_1 norm

$$||\mathbf{v}||_1 = \sum_{i=1}^{n} |x_i| \tag{4.5}$$

and the ℓ_∞ norm

$$||\mathbf{v}||_\infty = \max\{|x_1|, \ldots, |x_n|\}. \tag{4.6}$$

Bounding (4.7) and (4.8) below r results in ℓ_1 and ℓ_∞ balls, respectively. Thus, we quantify the mode-percentile misspecifications using the ℓ_1 and ℓ_∞ norms as

$$|m - m_T| + |p - p_T| \leq r \tag{4.7}$$

and

$$\max\{|m - m_T|, |p - p_T|\} \leq r, \tag{4.8}$$

respectively. In these cases, the extent of an expert's misspecification is readily interpretible. Using the ℓ_1 norm, the sum total of the (absolute) errors on the mode and percentile is at most r. In the ℓ_∞ case, the maximum misspecification on the mode and percentile is at most r. We show ℓ_1 and ℓ_∞ elicitation balls with their images in Figure 4.4.

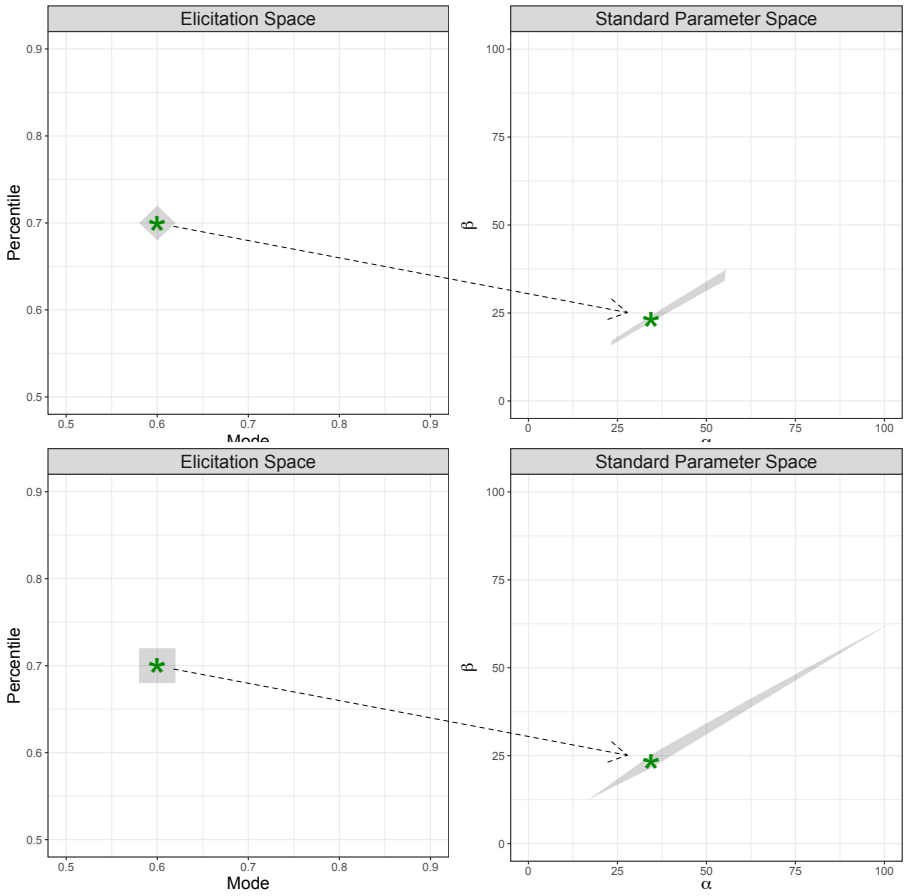


Figure 4.4: An ℓ_1 ball (top left) and an ℓ_∞ ball (bottom left) representing misspecifications of (m_T, p_T) mapped to images in standard parameter space.

4.2.2 Measuring Sensitivity

Each point in an image $\mathcal{I}$ results in a beta prior. Now that we have described the extent of the misspecification, we wish to measure how different the resulting priors can be. We assume $\pi_T(\theta) = \text{Beta}(\alpha_T, \beta_T)$ is a unique prior representing (m_T, p_T). We compare $\pi_T(\theta)$ to erroneous priors which result from misspecified summaries. A transformation (α, β) in an image $\mathcal{I}$ determines a beta prior $\text{Beta}(\alpha, \beta)$ prior with effective sample size $\text{ESS}(\alpha, \beta)$. We denote $\boldsymbol{\eta_T} = (\alpha_T, \beta_T)$. We compare erroneous ESS's to $\text{ESS}(\alpha_T, \beta_T)$ to

gauge the effect of the slight misspecifications on the resulting priors. To this end, we define our measure of sensitivity to be

$$\begin{aligned} \text{MESS}_{\boldsymbol{\eta}_T,r} &= \Big| \sup_{(\alpha,\beta)\in\mathcal{I}} \Big(\text{ESS}(\alpha,\beta) - \text{ESS}(\alpha_T,\beta_T)\Big)\Big| \\ &= \Big| \left(\sup_{(\alpha,\beta)\in\mathcal{I}} \text{ESS}(\alpha,\beta)\right) - \text{ESS}(\alpha_T,\beta_T)\Big| \\ &= \Big| \left(\sup_{(\alpha,\beta)\in\mathcal{I}} (\alpha+\beta)\right) - (\alpha_T+\beta_T)\Big| \end{aligned} \tag{4.9}$$

The MESS is a local measure of sensitivity. If the MESS is zero, then misspecifications do not cause any difference in the ESS of the resulting priors. Hence, a MESS of zero implies misspecifications do not impact prior strength. The larger the MESS, the more sensitive the method is to misspecification of (m_T, p_T). MESS is on an absolute, not a relative, scale, so it can be interpreted directly as the additional number of observations the misspecified prior would result in, in addition to those of the true prior.

4.2.3 Examples

In this section, we suppose that an expert's true beliefes about θ can be represented by a beta distribution with a mode of 0.4, and 95th percentile of 0.5, $\pi_T(\theta) = \text{Beta}(28.4, 42)$. We construct ℓ_2, ℓ_1 and ℓ_∞ balls, all with radius $r = 0.02$, representing misspecification around (0.4, 0.5), and we find their images in standard parameter space. The images resp-resent deviations from $\boldsymbol{\eta}_T = (28.4, 42)$. To illustrate our sensitivy measure, we calculate the MESS in the balls.

Figure 4.5 displays a ℓ_2 ball with its transformed image. The stars represent (m_T, p_T) and (α_T, β_T) in the elicited space and standard parameter space, respectively. The solid dots indicate the misspecified mode and percentile pair and resulting shape parameters yielding the MESS in the elicited space and standard parameter space, respectively. The MESS at at $(\alpha, \beta) = (55.8, 78.5)$ is 64. This corresponds to a misspecified mode and

95th percentile of 0.4142 and 0.4859, respectively. Here, the MESS of 64 results from $|(28.4 + 42) - (55.8 + 78.5)|$.

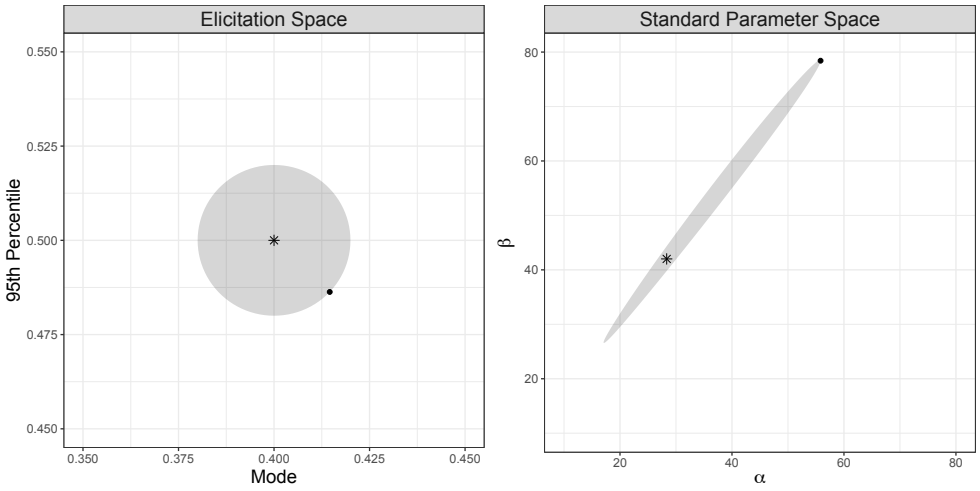


Figure 4.5: (Left) One elicitation ball with radius 0.02 centered at elicited mode and percentile, 0.4 and 0.5. (Right) Transformed ball in the standard parameter space.

In Figure 4.6, the MESS $= 38$ that occurs when the expert misspecifies her beliefs and gives a mode and 95th percentile of 0.4072 and 0.4872. This misspecificiation results in a Beta(44.4, 64.1) prior. It is not surprising that the MESS is less than that in Figure 4.5. The error structure determined by the ℓ_2 ball in (4.4) allows the mode and percentile, together, to have a larger distance from the true beliefs. Indeed, the ℓ_1 ball is a subset of the ℓ_2 ball, which is a subset of the ℓ_∞ ball.

An ℓ_∞ elicitation ball with the same center and radius as those in Figures 4.5 and 4.6 is shown in Figure 4.12. In this case, the MESS is 120, which results from a misspecified mode of 0.42 and 95th percentile of 0.48. These specifications result in a Beta(80.2, 110.4) prior.

We refer to the mode-percentile pair resulting in the MESS as the "worst" point. Notice that in Figures 4.5, 4.6, and 4.7, the worst point of misspecification is the point at the maximum distance from (m_T, p_T) in the fourth quadrant. This is the case for most, if not all, choices of centers for the ℓ_2 and ℓ_∞ balls. In our experience, the worst point

in ℓ_1 space varies but is always in the fourth quadrant. This is where the misspecified modes and percentiles grow closer together. The fact that the worst point occurs here is not surprising because the resulting priors are most informative here. We discuss this topic more in Section 4.3.1.

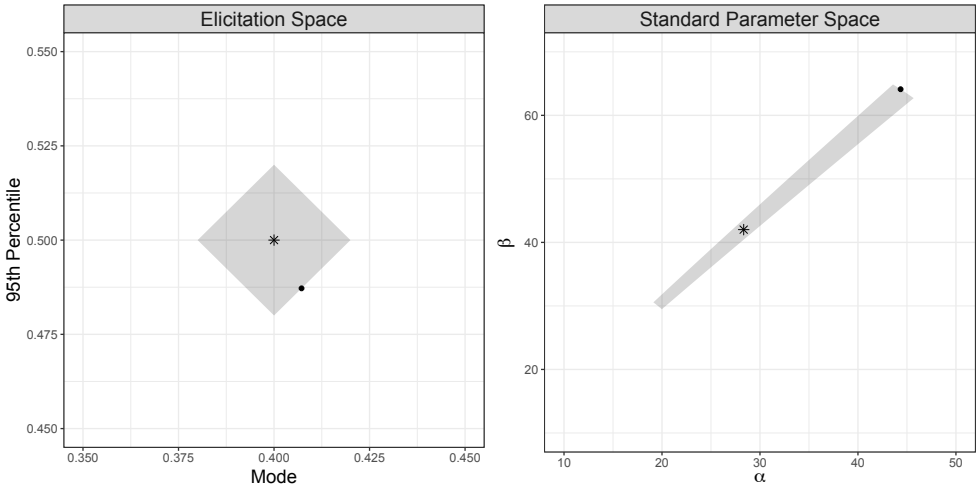


Figure 4.6: (Left) An ℓ_1 elicitation ball with radius 0.02 centered at elicited mode and percentile, 0.4 and 0.5. (Right) Transformed region in the standard parameter space.

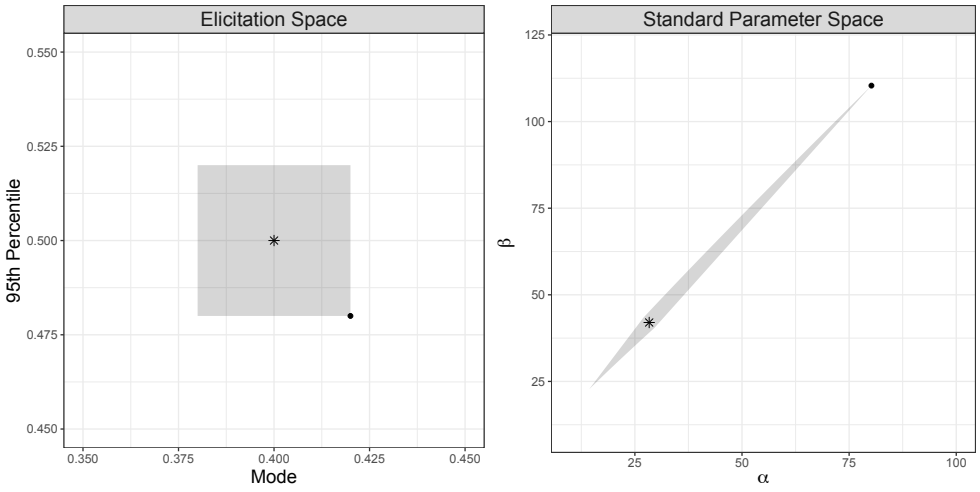


Figure 4.7: (Left) An ℓ_∞ elicitation ball with radius 0.02 centered at elicited mode and percentile, 0.4 and 0.5. (Right) Transformed ball in the (α, β) space.

4.3 A Tissot-Style Sensitivity Indicatrix

The previous discussion was limited to understanding the local distortion of the map of mode-percentile specifications into the standard parameter space at a single point in the space. To generalize it to the entire space, we consider many such balls like that in Figure 4.5 representing misspecification, each centered at a different point on a coarse mesh on the elicitation space, and transform each into the standard parameter space. We use the MESS to measure the sensitivity in each ball and denote the most sensitive regions with darker shading. We refer to the resulting plot in the elicitation space as a sensitivity indicatrix. This is reminiscent of Tissot's indicatrix, which has applications in geography and is used to measure local deformations in maps[2] .

In Figure 4.8, we tile the elicitation space with many ℓ_2 balls like that in Figure 4.5 to form a sensitivity indicatrix (left). The balls are centered at mode and percentile pairs and span possible true beliefs between 0.05 and 0.9. As before, we let the radius $r = 0.02$. The right plot in Figure 4.8 graphs the transformed images corresponding to the misspecification circles in the left plot. A larger diagonal ($y = x$ direction) stretch indicates a larger MESS. The images with the largest stretch are shaded the darkest, corresponding to the darkest-shaded elicitation balls.

The biggest MESS is 1161 which occurs when the expert misspecifies a her true beliefs of (0.45, 0.50) as (0.4646, 0.4863). This misspecification is a 0.03% increase of m_T and 0.03% decrease of p_T. In this case, using the elicited prior is equivalent to adding 1161 more prior observations to the model than the true prior would add! A sample size of 1161 could be over double what we would see in practice at times.

The overall pattern in the graphs indicates that as elicited modes and percentiles get closer together, the diagonal stretch in their transformations grows larger. Therefore, when the elicited mode and percentile values are close together, slight misspecifications of the expert's beliefs result in substantially large deviances in the corresponding priors.

[2] See the Wikipedia article on Tissot's indicatrix found at https://en.wikipedia.org/wiki/Tissot's_indicatrix.

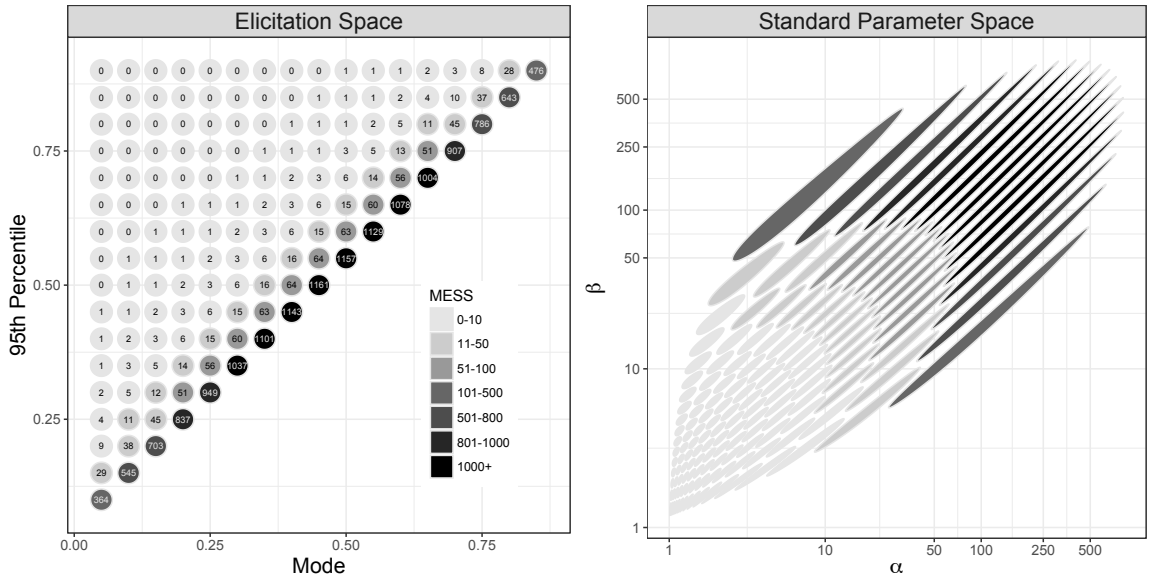


Figure 4.8: (Left) Many elicitation balls with radius 0.02 centered at elicited modes and percentiles. (Right) Transformed balls in the standard parameter space. The ball labeled 364 in the left image is the top-left most region in the graphic on the right. Note that the axes in the graph on the right are logarithmic.

How would one use the maps in practice? The maps are useful in informing the elicitation process itself. If the elicited values are in relatively stable regions of the parameter space, innaccuracies are not costly, and the elicitation process can be done relatively quickly. If the values are in sensitive regions of the parameter space, more caution is necessary to ensure that the elicited values are accurate, lest significant modeling errors occur.

We are interested, also, in the two "most extreme" points on the transformed balls. These points are the vertices corresponding to the major axes of the ellipses[3] in the standard parameter space. Comparison of these vertices can depict the maximum deviation in prior ESS's for a neighborhood around a pair of true shape parameters. Consider, for example, a scenario in which $(m_T, p_T) = (0.45, 0.50)$ yields $(\alpha, \beta) = (123.1, 150.2)$. In this elicitation ball, even small misspecifications could result in a prior more informative by over a thousand observations, or, on a relative scale, by 15 times. The corresponding transformed region $\mathcal{R}(\boldsymbol{\eta})$ is shown in Figure 4.9. The two most extreme points in this

[3] While the transformed balls appear to be ellipses, we have not yet proven this to be true.

transformed ball are (α_a, β_a) = (49.4, 63.5) and (α_b, β_b) = (666.6, 768.1). A misspecified mode-percentile pair of (0.4363, 0.5146) yields (α_a, β_a) which results in a prior with ESS 112.9. The transformation (α_b, β_b) results from mode-percentile pair (0.4646, 0.4863) and yields a prior ESS of 1434.7.

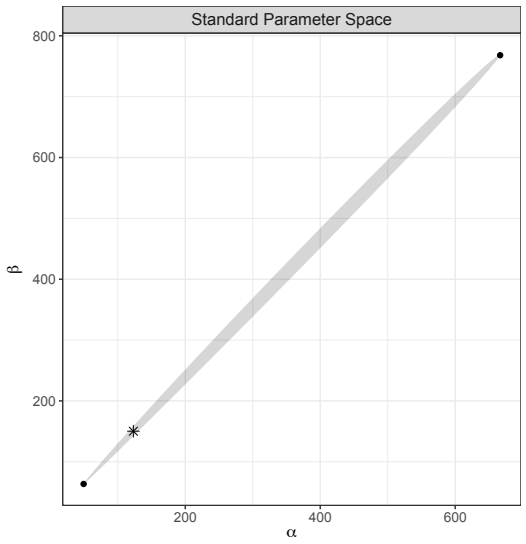


Figure 4.9: Transformed ball corresponding to the largest ESS. Solid dots represent extreme cases and the star indicates $\boldsymbol{\eta}_T$.

Figure 4.10 shows a sensitivity indicatrix tiled with ℓ_1 balls. In both Figure 4.8 and here, the biggest MESS occurs in the ball around $(0.45, 0.5)$. Here, the MESS is 481 which occurs when the expert misspecifies his beliefs and gives a mode of 0.4652 and 95th percentile of 0.4952. Notice, that the 95th percentile would round up to 0.5, but the decimals, which would otherwise be negligable, clearly make a difference in the resulting prior!

Lastly, a sensitivity indicatrix tiled with ℓ_∞ balls is shown in Figure 4.11. Again, the largest MESS is in the ball around $(m_T, p_T) = (0.45, 0.5)$. A misspecification of $(m, p) = (0.47, 0.48)$, less than 2% misspecifications, is equivalent to adding 6484 prior observations added to the model!

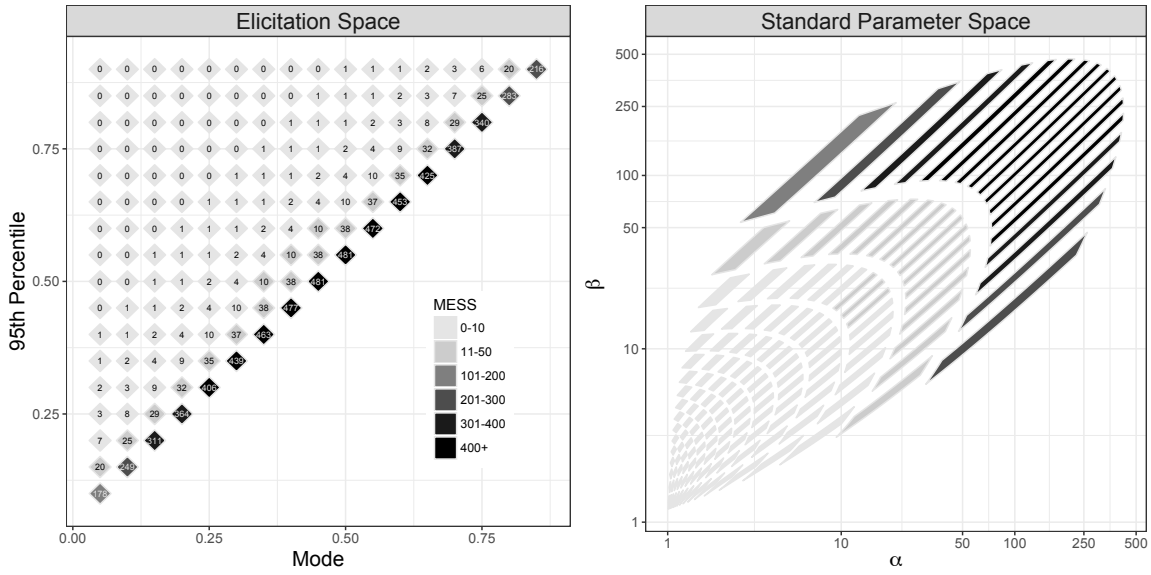


Figure 4.10: ℓ_1 elicitation balls with radius 0.02 centered at elicited modes and 95th percentiles; shaded according to MESS.

The ℓ_∞ ball around $(0.45, 0.5)$ is displayed in Figure 4.12. The image is so stretched that the region appears to be nearly a line. As in the previous cases, the worst misspecified mode and percentile pair is at the maximum distance in the fourth quadrant. The misspecification of $(0.47, 0.48)$ yields a Beta(3176.1, 3581.4) prior.

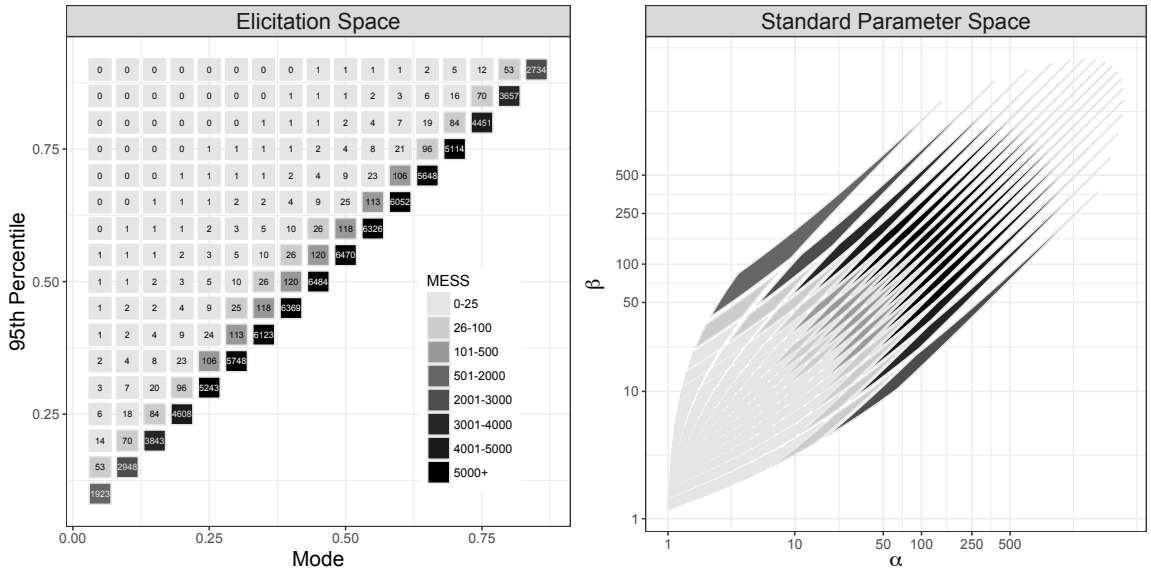


Figure 4.11: ℓ_∞ elicitation balls with radius 0.02 centered at elicited modes and 95th percentiles; shaded according to MESS.

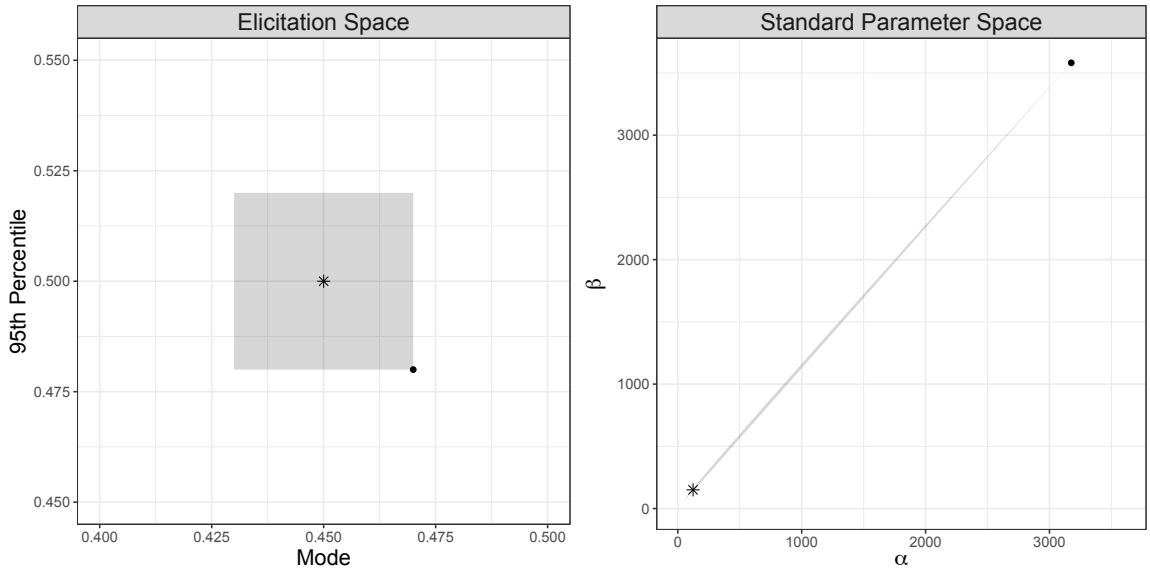


Figure 4.12: An ℓ_∞ ball around true beliefs (0.45, 0.5) and the resulting tranformed region. Stars represent truths and solid dots indicate misspecifications.

4.3.1 Lower Percentiles are Less MESSy

The fact that misspecifications on closely-spaced modes and percentiles yields large deviances in resulting priors is not surprising. As the distance between mode and percentile values decreases, the corresponding prior grows more informative. Figure 4.13 shows three prior distributions resulting from a specified mode of 0.4 and differing values of the 95th percentile. The most informative prior corresponds to a 95th percentile of 0.6, and the priors grow more diffuse as the value of the 95th percentile deviates more from the mode. Prior informativeness can be interpreted as the expert's confidence level in his specifications. To this end, if the expert is highly confident in his misspecified elicited values, the effect on the model is expected to be greater.

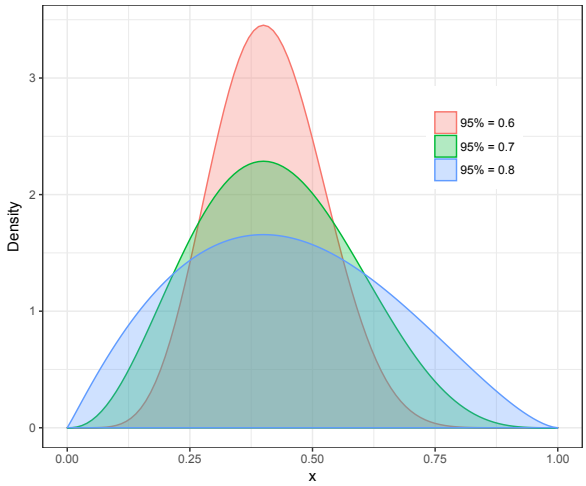


Figure 4.13: Beta priors resulting from elicited mode of 0.4 and differing values of the 95th percentile (95%).

In cases where modes and percentiles are close together, errors in elicitation not only have a large effect in terms of ESS but also, due to the nature of informative priors, the resulting prior has a large effect on the posterior analysis. Using lower percentiles results in a safer analysis. This is consistent with discussion in the literature, such as [41], which warn against the use of high percentiles in priors representing expert information. Experts tend toward over-confidence, and the 95th percentile, for example, gives only a small 5% chance that values exceed an elicited p. One way to adjust overly informative priors is to set the specified percentile value to a lower percentage, decreasing ξ.

We present sensitivity indicatrices with $\xi = 0.90$ and $\xi = 0.75$ in the left and right plots, respectively, of Figure 4.14. As expected, the impact of the expert's misspecifications lessens as ξ decreases. Here, the largest MESS is 706 and 197 when we set the percentile values to 90th and 75th percentile, respectively.

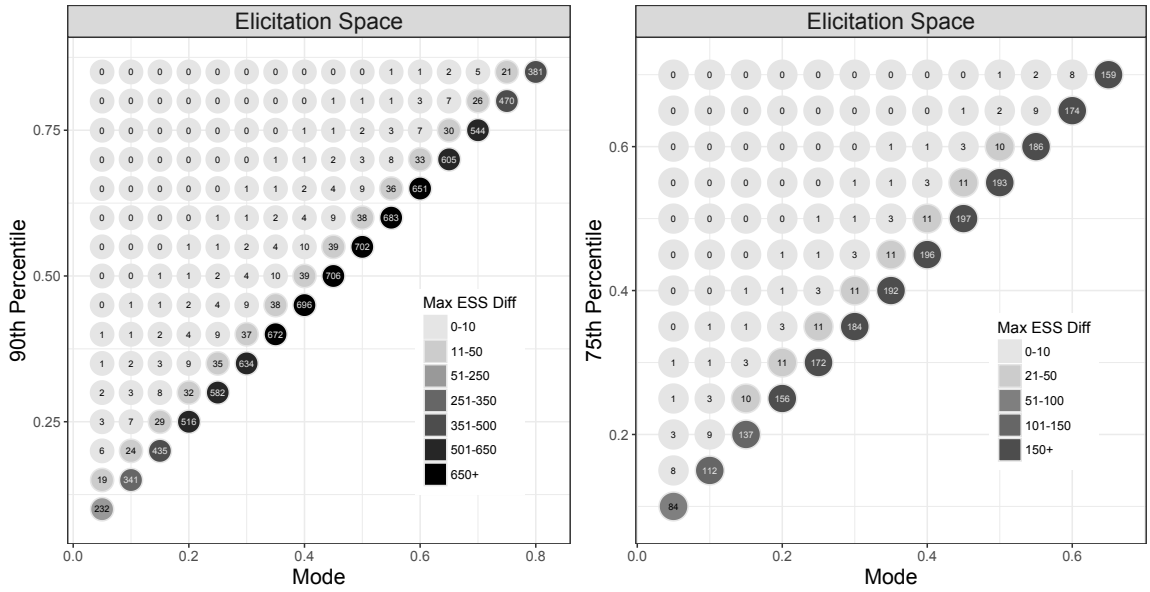


Figure 4.14: Elicitation balls with $\xi = 90$ (left) and $\xi = 75$ (right).

4.4 Conclusion

While it is well-known that information in prior elicitations can be misspecified, the literature lacks a systematic investigation into the consequences that this might have for the resulting prior. In this chapter, we created a mathematical framework with which to understand the effects of local misspecifications of bounded size. Furthermore, we presented a visual framework with which to undersand the local behavior more globally. In the case of the beta-Bernoulli model, the results mirror intuition and provide concrete numbers for the extent of the effect. The resulting maps can be used prescriptively to inform an elicitation procedure: if the expert specifies information in a sensitive region of the elicitation space, more care should be taken to ensure the accuracy of that information.

CHAPTER FIVE

Modelling Considerations in Network Meta-Analyses

In this chapter, we address modelling problems in network meta-analysis using simulated logistic regression data. Specifically, we consider prior specification on the between-trial standard deviation and baseline modelling.

5.1 Introduction

In this section, we provide a brief overview of meta-analysis and introduce the model we use for the remainder of the chapter.

5.1.1 A Brief Introduction to Meta-Analysis

Meta-analysis has been defined as the "statistical analysis of a collection of analytic results for the purpose of integrating the findings" [6]. The purpose is to combine results from different studies with common goals in order to reach an overarching conclusion. Meta analyses allow for combining trials with small sample sizes to improve influence. In effect, the trials borrow strength from one another to gain power. Further, results from meta-analyses can then be generalized to larger populations. DerSimonian and Laird [12] point out motivation for meta-analysis:

> Such analyses are becoming increasingly popular in medical research where information on efficacy of a treatment is available from a number of clinical studies with similar treatment protocols. If considered separately, any one study may be either too small or too limited in scope to come to unequivocable or generalizable conclusions about the effect of treatment. Combining the findings across such studies represents an attractive alternative to strengthen the evidence about the treatment efficacy.

In meta-anlayses, either a fixed-effect or random-effect model should be decided upon. Fixed effects models assume the studies are homogeneous; that is, there are no differences in underlying study populations, patient selection criteria, treatment protocol, etc. Patients assigned to the same treatment but in different studies are considered indistinguishable, at least up to included covariate values. That is, patients are assumed exchangeable across studies, as discussed in [51]. Random effects models treat study effects as drawn from a population of such effects. These models "borrow strength" across studies benefiting inference for both the population effect and individual study effects. Here we are assuming the studies are exchangeable.

Frequentist meta-analysis methods use a weighted average of point estimates (one from each study), with weights based on the standard errors of the estimates. In the fixed effect model, the weights are the inverse variances of the estimates. In the random-effect model, the between-study variation is incorporated into the weights. DerSimonian and Laird [12] provide a nice discussion of the latter. Bayesian meta-analysis methods are based on hierarchical models. Priors are elicited for parameters and corresponding hyperparameters. Priors become more diffuse as you move up in the hierarchy.

In addition to summarizing studies of the same treatment, the meta-analysis framework has been extended to provide indirect comparisons. Indirect comparisons are of interest when the treatments under comparison differ across studies, and we do not have a case where all studies compare all treatments. For example, suppose we have three studies. Study 1 might compare treatments A to B, Study 2 compares treatments B to C, and Study 3 compares treatments C to D. Network meta-analysis (NMA) provides methods for comparing A with C and D, as well as B with D. In this chapter, we focus on Bayesian random-effects models in network meta-analyses.

Suppose, for example, that treatment A is the primary reference. The general model for random effects, two-arm network meta-analysis is depicted in Figure 5.1. Here, y_{jk} is data from study j on treatment k. The data model is denoted by $\mathcal{D}$ and g is the link function.

Arrows respresent prior structures. The unknown parameters, $\boldsymbol{\theta}_{jk}$, include μ_{jb} and δ_{jb}, which represent baseline characteristics and treatment effects, respectively, of treatment b in study j. We use "after" in an alphabetically-ordered sense. In practice, it is often necessary to fix m and τ_μ, and we do so in subsequent sections. Also in subsequent sections, we consider various priors on $1/\sqrt{\tau_{\delta_{bk}}}$, including the uniform shown in the diagram.

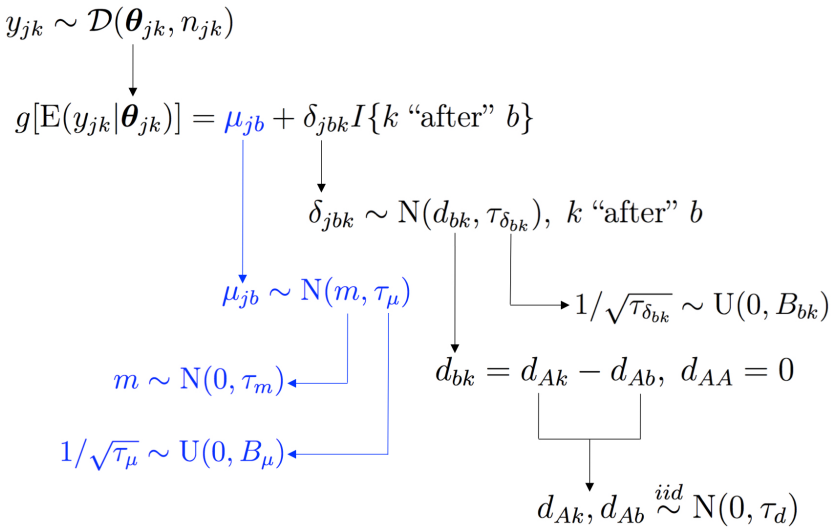


Figure 5.1: A diagram for network meta-analysis.

Network meta-analysis is also known as indirect and mixed treatment comparisons or multiple treatments meta-analysis. A major resource for network meta-analysis is a series of papers provided by the Decision Support Unit of the UK's National Institute for Health and Clinical Excellence [38]. The July 2013 issue of Medical Decision Making is devoted to evidence synthesis [15], [13],[14], [18], [16], [17], [1]. More details can also be found in Welton et al. [52].

5.1.2 Model for Simulations

In all simulations throughout the chapter, we generate data from N_S binary regression models, each with two arms. We assume 4 treatments are of interest: A, B, C, and D, treating A as the reference treatment. We give the meta-analysis design a star geometry,

meaning that each study compares A to either B, C, or D. For this reason, results for A versus B, C, or D are called direct comparisons because there is study data as direct evidence. Alternatively, B *vs.* C, C *vs.* D, and B *vs.* D are known as indirect comparisons and are obtained by subtracting corresponding direct comparisons. We assume the odds ratios are of interest, which we denote OR. The odds ratio comparing treatments B and C, for example, is obtained through:

$$OR_{BC} = OR_{AC} - OR_{AB}. \tag{5.1}$$

The relation in (5.1) requires a stringent consistency assumption, meaning that *if* we had a direct comparison for B *vs.* C, it would agree with the indirect comparison. For study $j = 1, ..., N_S$ and treatment $k =$ A, B, C, D, define

$$p_{jk} = \text{probability in study } j \text{ of a response to treatment } k,$$
$$r_{jk} = \text{count in study } j \text{ of response to treatment } k,$$

and

$$n_{jk} = \text{the number given treatment } k \text{ in study } j.$$

We simulate m data points from each of the N_S study models with

$$r_{jk} \sim \text{Bin}(p_{jk}, n_{jk}). \tag{5.2}$$

Let μ_{jb} and δ_{jbk} represent the baseline and relative treatment effects, respectively. The mean, d_{bk}, of the relative treatment effects is defined as the difference in effects of treat-

ments b and k. For a Bayesian analysis, we use the following model for all j and k:

$$\begin{aligned} \text{logit}(p_{jk}) &= \mu_{jb} + \delta_{jbk} I\{k \text{ after } b\}, \\ \mu_{jb} &\sim \text{N}(0, \tau_\mu), \\ \delta_{jbk} &\sim \text{N}(d_{bk}, \tau), \\ d_{bk} &= d_{Ak} - d_{Ab}, \\ d_{A\cdot} &\sim \text{N}(0, \tau_d), \\ \tau &= 1/\sigma^2, \end{aligned} \tag{5.3}$$

and

$$\sigma \sim g(\cdot, \cdot),$$

where the normal priors are defined in terms of the mean and precision, and $g(\cdot, \cdot)$ is a prior with support a subset of the positive reals. We treat the treatment effects as random. The odds ratios (OR's) are of interest. The difference in treatment effects $d_{bk} = d_{Ak} - d_{Ab}$ is equal to the log odds ratio $\log(OR_{bk})$. Similarly, the direct effects are $d_{Au} = \log(OR_{Au})$, $u = k, b$. Thus, the OR's have diffuse log-normal induced priors, which we discuss in detail in Section 5.4. The baseline effect, μ_{jb}, includes characteristics from each study not accounted for in the treatment, such as age, race, and disease stage.

In Section 2, we use a basic model to account for baseline effects, and we focus on specifying $g(\cdot, \cdot)$ for the between-trial standard deviation. We investigate an alternative baseline model in Section 3.

5.2 Between-Trial Heterogeneity

Studies used for meta-analyses may differ in terms of design, methodology, and sampling error due to differing sample sizes. These differences among studies can produce between-trial heterogeneity. Potential sources of heterogeneity include, for example, difference in treatment regimens, patient eligibility criteria, baseline disease severity, and outcomes [43] . In our model, we capture this heterogeneity in the between-trial standard

deviation σ. In this section, we detail some problems that can occur in modeling σ. We compare priors on σ through a simulation study in Section 2.1, and we compare parameterization of specified priors on σ in Section 2.2.

Some priors commonly used to model the between-study heterogeneity σ include the half-normal, log-normal, half-student, uniform, and gamma denoted:

$$\begin{aligned} \sigma &\sim \text{Half-N}(0, \psi), \\ \sigma &\sim \text{Log-N}(0, \rho), \\ \sigma &\sim \text{Half-Student}(\nu, \eta), \\ \sigma &\sim \text{U}(0, B), \end{aligned} \tag{5.4}$$

and

$$\tau \sim \text{Gamma}(\xi, \xi),$$

respectively. The half-normal prior is the positive half of a N$(0, \psi)$ distribution. That is, if $u \sim$ N$(0, \psi)$, then $|u|$ is distributed Half-N$(0, \psi)$. Similarly, the half-student prior is the positive half of a student-t distribution with density

$$\pi(\sigma) \propto \left[1 + \frac{1}{\nu}\left(\frac{\sigma}{\eta}\right)^2\right]^{-(\nu+1)/2}$$

where η is a scale parameter and ν is degrees of freedom. The log-normal is a N$(0, \rho)$ exponentiated. Finally, the gamma prior is in terms of equal shape and scale parameters and formulated in the same way as previous chapters. Often, prior parameters are chosen so as to yield "relatively non-informative" priors. For example, one may take B "large" or ξ "small". Such choices can be problematic, as we shall see.

Figure 5.2 shows examples of the first four priors listed above. The middle column contains graphs of the prior density on σ, and the left and right columns are the resulting induced priors on the precision, τ, and the variance. The first through fourth rows correspond, respectively, to priors: Half-N(0,1), Log-N(0, 1), Half-Student(10, 1), and U(0, 2).

Plots for the gamma prior listed last in (5.4) are not included above. Gelman [24] notes problems with this prior. Choosing small ξ, say $\xi = 0.001$, induces extremely diffuse priors on σ and the variance. Such highly diffuse priors are known to cause problems with convergence because extreme values of σ can be sampled. Further, the corresponding prior on τ is highly concentrated near the origin implying extreme probability that τ is close to zero. In effect, the gamma prior on τ is highly informative while the resulting induced prior on σ is overly diffuse.

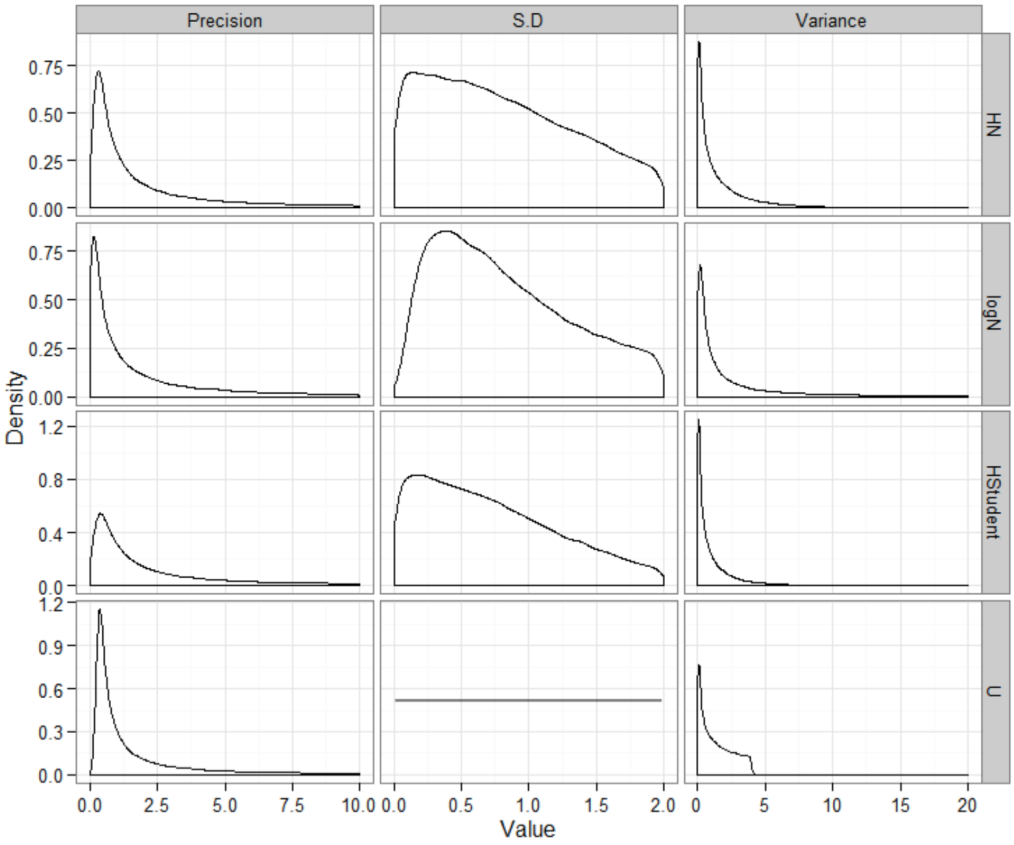


Figure 5.2: Density plots on the between-trial standard deviation (S.D.) with induced priors on the precision and variance.

As with any choice of prior distribution, prior-to-posterior sensitivity analyses are recommended. Consider choosing the upper bound, B, for the uniform prior in (5.4). Values of B that are too large can cause convergence problems, but small values of B can

overly influence the posterior results. For example, suppose we are interested in a 95% credible interval on a particular odds ratio. We can plot the credible interval against various values of B, as shown in Figure 5.3 for an hypothetical study. The credible interval widths stabilize when $B = 1.5$. This suggests that a value of B around 3 may be a good choice.

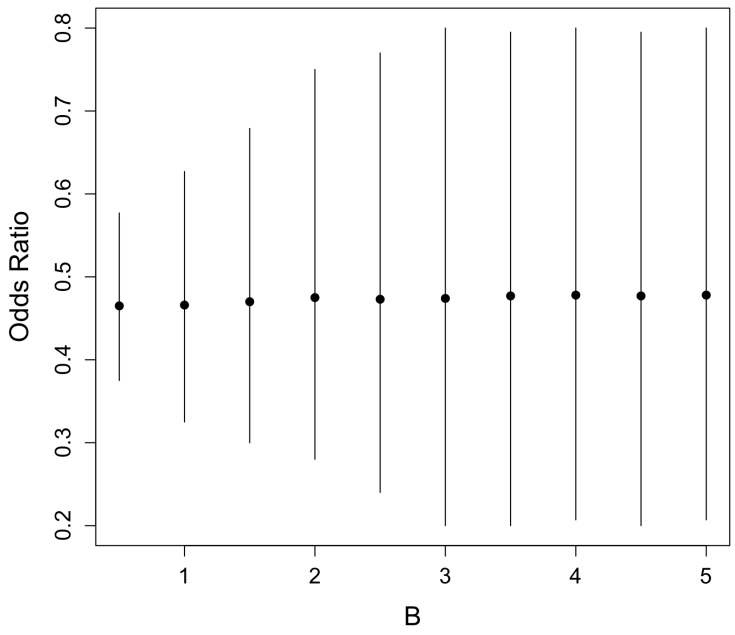


Figure 5.3: Credible intervals for an odds ratio from an hypothetical NMA varying with the upper bound, B, in the uniform prior on σ.

It is also helpful to keep in mind practical values of the parameters and corresponding implications of the prior specification. Spiegelhalter et al. [49] provide more detail on this point. For example, consider our model in (5.3) where the log-odds ratios δ_{jbk} are assumed normal with mean d_{bk} and standard deviation σ. Then, 95% of the δ_{jbk}'s will be in the interval $d_{bk} \pm 1.96\sigma$. Furthermore, the ratio of 97.5th to 2.5th percentiles of the odds ratios $(\exp(\delta_{jbk}))$ is $\exp(3.92\sigma)$. Therefore, if we believe that the odds ratios from the various studies differ by no more than a ratio of 15, we can set $\exp(3.92\sigma)$ equal to 15. This suggests setting $\sigma = 0.7$.

5.2.1 *Prior Specification on Between-Trial Standard Deviation*

In this section we investigate the sensitivity of posterior results to the specification of priors on the between-trail standard deviation, σ, in network meta-analyses. We consider uniform, log-normal, and half-normal priors for $g(\cdot, \cdot)$ in (5.3). Also, we consider an informative gamma prior on τ which induces a diffuse inverse-gamma on σ^2. We vary the numbers of studies and study sample sizes. We generate 500 different data sets for each case and compare average posterior summaries.

Below are results from simulations investigating the consequences of different choices for $g(\cdot, \cdot)$ on the random effects standard deviation, σ, in network meta-analyses with $N_S = 9$ and $N_S = 18$ studies. The 500 datasets are simulated using the true value for μ simulated from a N(–1, 0.4) distribution. The sample sizes n_S and n_L vary for each study, where n_S and n_L denote small and large sample sizes ranging from [50,100] and [150,250], respectively. We chose these sample sizes by generating n_S and n_L from a U(50, 100) and U(150, 250), respectively, for each study. We set $\tau_d = \tau_\mu = 0.0001$ which results in extremely diffuse priors on the μ's and d's, especially for a logistic model. We consider the following priors in the simulation: Gamma(0.001, 0.001) on τ, and U(0, 5), Log-N(0, 0.503^2), and Half-N(0, 0.3145) on σ. The log-normal and half-normal priors were chosen to correspond to standard deviations of 0.8 and 2.4, respectively, corresponding to real examples we recently worked on with industry colleagues. The gamma on the precision is extremely concentrated around zero inducing a prior on σ that is overly diffuse. The gamma prior in Figure 5.4 is the prior on σ induced by a Gamma(0.001, 0.001) prior on τ. The density is approximately uniform on $(0, 10^{100})$.

Posterior results are found in WinBUGS using 30,000 MCMC iterations following a burn-in of 10,000 iterations.[1] Some cases did not converge after 30,000 iterations. Therefore, in these cases, sample sizes were increased at random to 1000 which allowed the chains to converge. This could cause bias toward the study(s) whose sample size was

[1] This study was motivated by questions from biopharmaceutical industry colleagues. They needed the iterations to be kept at 30,000 in order to test a new software package.

increased to 1000, especially in the cases where sample sizes for all other studies were between 50 and 100.

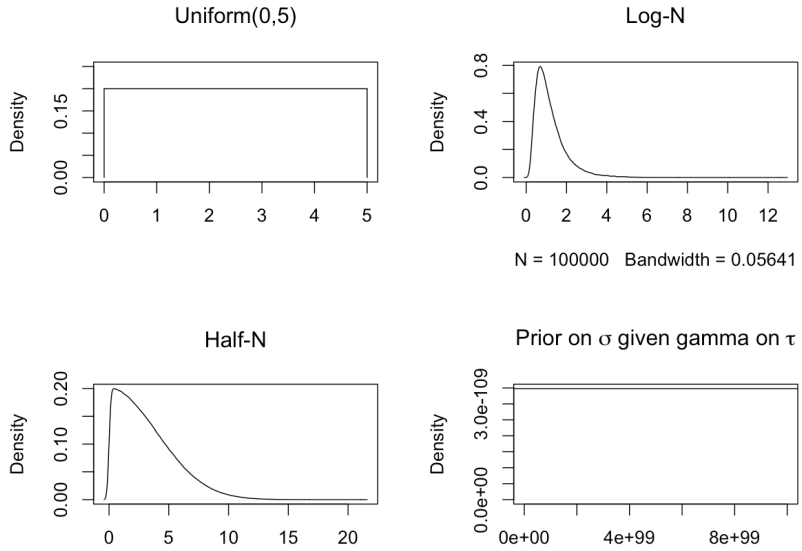


Figure 5.4: Density plots of the four priors used for σ in simulations.

The chains that used the Gamma(0.001, 0.001) prior did not converge in OpenBUGS. Therefore, convergence was attempted in JAGS but was also not reached. Indeed, these chains produced posterior means for odds ratios of up to 10^{70}. The non-convergence and large posterior means provide more evidence concerning our warning on gamma priors in Section 5.2 above. The prior induced on σ could be overly diffuse.

Figure 5.5 exhibits the poor behavior of a chain that used a Gamma(0.001, 0.001) prior. The chain is for OR_{BD} in the 9-study case with small samples sizes. Notice that many iterations sample extreme values, as high as 200,000. As we will see, this extreme behavior is a consequence of the combination of the Gamma(0.001, 0.001) prior on τ and the overly diffuse normal prior on the treatment effects.

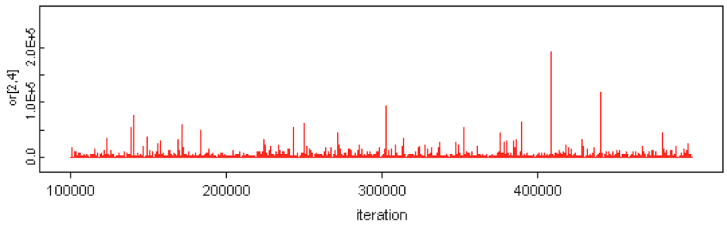


Figure 5.5: History plot of a chain for OR_{AB} given the Gamma(0.001, 0.001) prior on τ.

The average of 500 posterior means, coverages, and CI lengths are displayed in Table 5.1 below. In all cases, the average posterior means over-estimate the true OR's. As we will see, this is a consequence of the overly diffuse normal on the treatment effects. We discuss this in Section 5.4. For the 18-studies case with large sample sizes, the posterior means using the uniform prior are closest to the true values. The log-normal prior produces posterior means closest to the true values in the 6-studies case with small sample sizes.

As expected, the average credible interval lengths decrease as the sample data increases for every prior. Overall, the decrease in interval length leads to decreases in coverages. As the interval shrinks, there is less variability to capture the true OR value. There is no one prior that substantially out-performed the others. Credible interval lengths and coverages were comparable across all priors. The uniform and inverse-gamma priors provided the best average posterior means. However, the inverse-gamma prior can produce outlying means. Also, the uniform prior is simple in its interpretation and implementation.

Note that the choice of μ for these datasets yields well-behaved logistic data. Therefore, all priors seem to perform well. In the next section, we consider these priors in logistic datasets with fewer successes and small sample sizes.

5.2.2 Prior Performance in Data with Few Successes

In this section we perform the same simulation as detailed in Section 3.1, but we make two changes to the 500 simulated data sets. We simulate the data using a mean $\mu \sim$ N(–3, 0.25) yielding logistic data with much fewer successes. Also, we consider meta-

analyses with $N_S = 6$ and $N_S = 9$ studies and only small sample sizes, which is a data reduction compared to the above simulation study. Overall, this data is not as well-behaved as the data in the previous section.

This data immediately led to convergence issues in OpenBUGS. After one million iterations (with burn-in's of 250,000), there was non-convergence with every prior. This forced an increase in the precisions τ_d and τ_μ. Non-convergence was still present with $\tau_d = \tau_\mu = 0.01$. Therefore, in these simulations, we use $\tau_d = \tau_\mu = 0.1$. With this precision, convergence was still difficult to attain. OpenBUGS produced errors when using the half-normal prior with both 6 and 9 studies and the uniform prior with 6 studies, and convergence was not reached in JAGS. As before, convergence was not reached with chains that used the gamma prior. The simulation using the log-normal with 6 studies converged after 850 thousand iterations (with a burn-in of 100 thousand) in OpenBUGS. We used 1.5 million iterations with a burn-in of 150 thousand and thin of 25 in JAGS for the log-normal with 9 studies, and 2.5 million iterations with a burn-in of 350 thousand and thin of 50 for the uniform with 9 studies. These simulations were not time-efficient; some taking up to 20 minutes for one data-set. Table 5.2 shows averages of the 500 posterior means, 95% credible interval lengths, and coverages.

The log-normal was the only prior that produced convergent chains using 6 studies. Convergence was reached only with log-normal and uniform priors in the 9-studies cases. The log-normal performed better than the uniform prior. Specifically, the posterior means were closer to the true values and the CI lengths were smaller in the log-normal case.

While convergence was not reached with the Gamma(0.001, 0.001) prior, the values sampled were not nearly as large as in Section 5.2.1. Recall that, before, we used $\tau_d = 0.0001$ so that the treatment effects were given a normal prior with 0 mean and a precision of 0.0001. Here, we use $\tau_d = 0.1$. Indeed, posterior means for the odds ratios were as large as 14,000, which is extreme, but much less than the 10^{70} that we saw before. We discuss the choice of prior on the treatment effects in Section 5.4.

Table 5.1: Posterior simulation results for "well-behaved" data sets.

True Values		$\sigma \sim$ Unif				$\sigma \sim$ Inv-Gamma				$\sigma \sim$ Log-N				$\sigma \sim$ Half-N			
		$N = 9$		$N = 18$		$N = 9$		$N = 18$		$N = 9$		$N = 18$		$N = 9$		$N = 18$	
		n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L	n_S	n_L
$OR_{AB} = 1.221$	Mean	1.601	1.481	1.371	1.316	NA	NA	NA	NA	1.465	1.456	1.359	1.316	1.539	1.445	1.359	1.316
	Coverage	0.972	0.958	0.954	0.950	NA	NA	NA	NA	0.986	0.976	0.970	0.970	0.960	0.980	0.960	0.960
	Length	3.357	2.582	1.581	1.315	NA	NA	NA	NA	2.902	2.498	1.603	1.335	3.129	2.528	1.561	1.304
$OR_{AC} = 2.014$	Mean	2.606	2.364	2.223	2.126	NA	NA	NA	NA	2.532	2.309	2.244	2.206	2.597	2.427	2.241	2.203
	Coverage	0.974	0.966	0.966	0.956	NA	NA	NA	NA	0.962	0.978	0.954	0.966	0.974	0.966	0.934	0.958
	Length	5.452	4.103	2.509	2.100	NA	NA	NA	NA	5.075	3.960	2.595	2.230	5.258	4.245	2.520	2.180
$OR_{AD} = 1.649$	Mean	2.110	1.953	1.851	1.737	NA	NA	NA	NA	2.005	1.913	1.806	1.749	2.094	1.909	1.805	1.748
	Coverage	0.964	0.974	0.958	0.962	NA	NA	NA	NA	0.984	0.990	0.972	0.978	0.964	0.974	0.956	0.966
	Length	4.303	3.401	2.091	1.723	NA	NA	NA	NA	3.929	3.279	1.979	1.736	4.227	3.378	1.912	1.695
$OR_{BC} = 1.649$	Mean	3.251	2.623	1.896	1.823	NA	NA	NA	NA	2.659	2.164	1.941	1.885	3.929	2.542	1.932	1.880
	Coverage	0.978	0.966	0.956	0.958	NA	NA	NA	NA	0.976	0.992	0.974	0.978	0.978	0.972	0.964	0.970
	Length	8.210	5.733	3.114	2.589	NA	NA	NA	NA	7.729	5.362	3.254	2.749	8.126	6.014	3.160	2.688
$OR_{BD} = 1.350$	Mean	2.404	2.024	1.579	1.491	NA	NA	NA	NA	2.068	1.775	1.547	1.501	2.611	1.961	1.541	1.497
	Coverage	0.976	0.978	0.954	0.948	NA	NA	NA	NA	0.988	0.992	0.972	0.966	0.972	0.982	0.964	0.958
	Length	6.477	4.774	2.591	2.128	NA	NA	NA	NA	6.047	4.451	2.523	2.168	6.525	4.769	2.449	2.119
$OR_{CD} = 0.819$	Mean	1.649	1.238	0.963	0.913	NA	NA	NA	NA	1.203	1.109	0.946	0.890	3.334	1.277	0.942	0.889
	Coverage	0.968	0.968	0.964	0.962	NA	NA	NA	NA	0.978	0.986	0.954	0.956	0.978	0.964	0.942	0.950
	Length	3.878	2.940	1.567	1.304	NA	NA	NA	NA	3.418	2.748	1.530	1.273	3.713	2.851	1.482	1.244
$\sigma = 0.500$	Mean	0.663	0.621	0.538	0.542	NA	NA	NA	NA	0.679	0.640	0.576	0.558	0.637	0.622	0.542	0.540
	Coverage	0.944	0.944	0.944	0.952	NA	NA	NA	NA	0.930	0.956	0.962	0.936	0.954	0.948	0.952	0.934
	Length	1.419	1.122	0.692	0.522	NA	NA	NA	NA	1.018	0.888	0.586	0.491	1.371	1.113	0.638	0.510

All of the average posterior means are positively biased. As we will see, this is due to the induced prior on the odds ratios given the diffuse normal priors on the treatment effects. We discuss this further in Section 5.4. Many posterior means are quite a bit larger than the truth which is cause for concern. Of particular concern, the posterior means estimate OR_{CD} to be greater than 1 in all cases, but OR_{CD} is truly less than 1. This result, especially, could result in an incorrect decision in practice.

Further, the credible interval lengths are impractically large. As expected, the lengths are extremely large in the 6-study cases and for the indirect comparisons. In fact, these may be too large to make any real conclusions in some trials. Suprisingly, for OR_{AD}, the credible interval lengths are larger in the 9-study cases than in the 6-study cases. To investigate whether this is due to a few outlying lengths, we have included the median of the lengths (labeled "Median Length") in Table 5.2. While the difference between the average and median of the 500 lengths suggests a skewed distribution of lengths, the median lengths are still larger in the 9-study cases. This can be seen in Figure 5.6 as well.

Average posterior medians are comparable across all priors. This implies that extreme values are being sampled and skewing the posterior causing the larger values of and variability between the means. The posterior median is robust to the outliers, whereas the mean is skewed with the distribution. Because there is evidence of skewed posteriors, using the posterior median as a point estimate rather than the mean is recommended. As expected, medians using 6 studies diverge more from the truth than those using 9 studies.

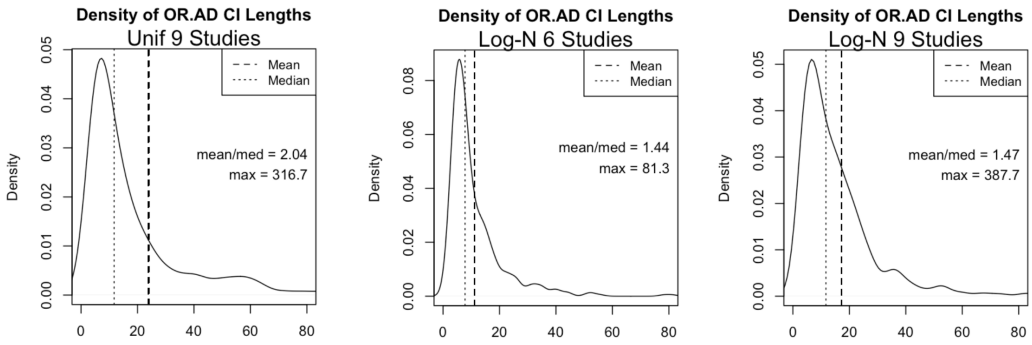


Figure 5.6: Densities of 500 posterior credible interval (CI) lengths for OR_{AD}. The ratios between mean and median and the maximum CI lengths are specified.

Table 5.2: Posterior simulation results for datasets with few successes and small samples.

True Values		$\sigma \sim$ Unif		$\tau \sim$ Gamma		$\sigma \sim$ Log-N		$\sigma \sim$ Half-N	
		$N_S = 6$	$N_S = 9$	$N_S = 6$	$N_S = 9$	$N_S = 6$	$N_S = 9$	$N_S = 6$	$N_S = 9$
$OR_{AB} = 1.221$	Mean	NA	1.87	NA	NA	2.05	1.70	NA	NA
	Median	NA	1.37	NA	NA	1.34	1.32	NA	NA
	Coverage	NA	0.994	NA	NA	0.992	0.986	NA	NA
	Length	NA	5.86	NA	NA	7.66	4.91	NA	NA
$OR_{AC} = 2.014$	Mean	NA	3.11	NA	NA	3.65	2.83	NA	NA
	Median	NA	2.26	NA	NA	2.40	2.20	NA	NA
	Coverage	NA	0.988	NA	NA	0.994	0.986	NA	NA
	Length	NA	9.78	NA	NA	13.57	8.14	NA	NA
$OR_{AD} = 1.649$	Mean	NA	5.95	NA	NA	2.90	4.07	NA	NA
	Median	NA	2.26	NA	NA	1.85	2.24	NA	NA
	Coverage	NA	0.988	NA	NA	0.990	0.994	NA	NA
	Length	NA	**28.90**	NA	NA	11.17	**17.17**	NA	NA
	Median Length	NA	11.72	NA	NA	7.78	11.70	NA	NA
$OR_{BC} = 1.649$	Mean	NA	2.23	NA	NA	**13.95**	2.14	NA	NA
	Median	NA	1.66	NA	NA	2.97	1.69	NA	NA
	Coverage	NA	0.992	NA	NA	0.994	0.980	NA	NA
	Length	NA	6.75	NA	NA	**64.22**	6.01	NA	NA
$OR_{BD} = 1.350$	Mean	NA	2.24	NA	NA	9.90	3.16	NA	NA
	Median	NA	1.63	NA	NA	2.20	1.74	NA	NA
	Coverage	NA	0.990	NA	NA	0.992	0.982	NA	NA
	Length	NA	16.30	NA	NA	**44.81**	13.24	NA	NA
$OR_{CD} = 0.819$	Mean	NA	2.14	NA	NA	3.15	1.67	NA	NA
	Median	NA	0.95	NA	NA	1.11	1.01	NA	NA
	Coverage	NA	0.984	NA	NA	0.994	0.984	NA	NA
	Length	NA	8.18	NA	NA	13.98	6.61	NA	NA
$\sigma = 0.500$	Mean	NA	0.86	NA	NA	0.92	0.83	NA	NA
	Median	NA	0.75	NA	NA	0.83	0.76	NA	NA
	Coverage	NA	0.958	NA	NA	0.970	0.964	NA	NA
	Length	NA	2.23	NA	NA	1.76	1.42	NA	NA

Table 5.3 displays results using the same model but with different parameter values. Specifically, here $\delta_{AB} = \delta_{AC} = \delta_{AD} = 0.9$ and $\sigma = 0.3$. This yields odds ratios for all direct and indirect comparisons of 2.46 and 1, respectively. The simulations using the half-normal prior and uniform prior with 6 studies did not converge in JAGS and errored in OpenBUGS. The gamma cases did not converge as the chain sampled extreme values similar to those discussed above.

Remaining simulations with 9 studies were run in JAGS. For the uniform prior, we used 2.5 million iterations with 150 thousand as burn-in and thinned every 50. We ran 1.5 million iterations for the log-normal prior with a burn-in of 150 thousand and thin of 25. We ran 850 thousand iterations in OpenBUGS for the log-normal prior with 6 studies. This included a burn-in of 100 thousand and no thinning. The Gelman-Rubin statistic, smooth posterior densities, and history plots fall provided evidence of converegence in all cases. Further, autocorrelations plots converge to zero after the necessary thinning.

With this data, the uniform prior resulted in posterior means closest to the true values. This is similar to the results in Table 5.2. However, here, the credible interval length between the two 9-study cases are comparable. In Table 5.2, the credible interval lengths using the uniform prior were larger than those with the log-normal prior.

5.3 Impact of the Parameterization of Priors

In this section, we simulate data sets to test the effect of the parameterization in the uniform and gamma priors.

5.3.1 Uniform Upper Bound

We first consider a uniform prior on the between-study standard deviation σ so that

$$\sigma \sim \mathrm{U}(0, B).$$

We fix the lower bound at 0 and investigate the effect of changes in the upper bound B on the posterior results. To do so, we simulate 100 data sets, each with 9 studies with sample sizes between 50 and 100, following the model in (5.3) with $\sigma = 0.4$. True values of the other variables and the posterior results are shown in Table 5.4. The true value of μ was simulated from a normal distribution with mean –2 and standard deviation of 0.25. We chose this data because datasets from the previous sections did not converge in JAGS or OpenBUGS using$B = 100$. Also, we restricted the data to 9 studies because datasets with 6 studies did not converge.

For the $B = 2$ case, we used 3.5 million iterations in JAGS with a burn-in of 250 thousand and thinned every 30. For $B = 5$, we also used a thin of 30 with 1.5 million iterations and a burn-in of 150 thousand. For the $B = 50$ and $B = 100$ cases, we ran 3.5 million iterations in JAGS with a burn-in of 250 thousand and thin of 50. The results displayed in Table 5.4 are the averages of 100 posterior means, coverages, and 95% credible lengths.

The simulation results imply that there is no real difference between the uniform priors with upper bounds of 10, 50 and 100. This is particularly apparent in the $B = 50$ versus $B = 100$ cases, as the simulations yield the same results. The slight differences seen in the $B = 10$ case can be attributed to Monte Carlo error in the simulations. All results are biased in the positive direction. It is concerning that the posterior means for OR_{CD} are above 1 while the truth is below 1. This could yield opposing conclusions in studies based upon odds ratios. Particular care should be taken when using the uniform prior in logistic data sets with little successess such as these.

Table 5.3: More posterior results for simulations with few successes and small samples sizes.

True Values		$\sigma \sim$ Unif		$\tau \sim$ Gamma		$\sigma \sim$ Log-N		$\sigma \sim$ Half-N	
		$N_S = 6$	$N_S = 9$	$N_S = 6$	$N_S = 9$	$N_S = 6$	$N_S = 9$	$N_S = 6$	$N_S = 9$
$OR_{AB} = 2.46$	Mean	NA	3.18	NA	NA	3.86	3.11	NA	NA
	Median	NA	2.57	NA	NA	2.70	2.55	NA	NA
	Coverage	NA	0.99	NA	NA	1.00	0.988	NA	NA
	Length	NA	8.13	NA	NA	13.33	7.97	NA	NA
$OR_{AC} = 2.46$	Mean	NA	3.35	NA	NA	3.95	3.36	NA	NA
	Median	NA	2.69	NA	NA	2.71	2.71	NA	NA
	Coverage	NA	1.00	NA	NA	0.996	0.984	NA	NA
	Length	NA	8.79	NA	NA	13.99	8.94	NA	NA
$OR_{AD} = 2.46$	Mean	NA	2.21	NA	NA	4.01	1.97	NA	NA
	Median	NA	1.30	NA	NA	2.69	1.29	NA	NA
	Coverage	NA	0.864	NA	NA	0.998	0.876	NA	NA
	Length	NA	7.72	NA	NA	14.62	7.35	NA	NA
$OR_{BC} = 1$	Mean	NA	3.00	NA	NA	3.08	3.16	NA	NA
	Median	NA	2.46	NA	NA	1.31	2.60	NA	NA
	Coverage	NA	0.786	NA	NA	0.998	0.752	NA	NA
	Length	NA	7.51	NA	NA	13.83	8.06	NA	NA
$OR_{BD} = 1$	Mean	NA	1.97	NA	NA	3.06	1.82	NA	NA
	Median	NA	1.19	NA	NA	1.27	1.21	NA	NA
	Coverage	NA	0.984	NA	NA	1.00	0.992	NA	NA
	Length	NA	6.77	NA	NA	13.70	6.61	NA	NA
$OR_{CD} = 1$	Mean	NA	1.77	NA	NA	3.07	1.79	NA	NA
	Median	NA	1.10	NA	NA	1.31	1.20	NA	NA
	Coverage	NA	0.996	NA	NA	0.998	0.986	NA	NA
	Length	NA	6.07	NA	NA	13.74	6.47	NA	NA
$\sigma = 0.300$	Mean	NA	0.63	NA	NA	0.84	0.72	NA	NA
	Median	NA	0.53	NA	NA	1.62	0.67	NA	NA
	Coverage	NA	0.978	NA	NA	0.816	0.832	NA	NA
	Length	NA	1.77	NA	NA	1.62	1.25	NA	NA

5.3.2 Gamma Parameters

In this section, we give the precision τ a gamma prior and use simulated data sets to investigate the effect of the gamma parameterization on the posterior results. We have already noted problems with the Gamma(0.001, 0.001) which induces an overly diffuse prior on σ. This implies that σ might take on some impractical extreme values and cause outliers in the posterior means.

For completeness, we consider $\tau \sim$ Gamma(0.001, 0.001) again. We also use three other gamma priors with shape and scale parameters, ξ, of 0.5, 0.1, and 0.01. These priors have means near 0 with variances of 0.125, 0.001, and 1×10^{-6}, respectively. The Gamma(0.001, 0.001) has variance 1×10^{-9}. The smaller the shape and rate parameters, the more concentrated near 0 the prior on τ, and the induced prior on σ is more diffuse. The Gamma(0.5, 0.5) and Gamma(0.1,0.1) priors on τ induce priors on σ with means of approximately 9.31 and 2.7×10^{18} and variances of approximately 20,600 and 2.4×10^{41}, respectively. The resulting induced priors on σ from both the Gamma(0.01,0.01) and Gamma(0.001, 0.001) priors have means of infinity. The Gamma(0.001, 0.001) prior induces a density on σ that is approximately uniform on 0 to 10^{100}, and the Gamma(0.01, 0.01) induces a density on σ that is approximately uniform on 0 to 10^{18}. (All means, variances, and densities computed numerically in `R`). While prior variances are extremely large, representing large uncertainty, prior means are also extremely large. This is evidence that these priors exhibit extreme positive bias.

Using 100 simulated data sets with $\mu \sim$ N(–3, 0.25), $\sigma = 0.5$, $d_{AB} = 0.1, d_{AC} = 0.85$, and $d_{AD} = 0.4$, we obtain the average of 100 posterior means, coverages, and 95% credible interval lengths using the various gamma priors on τ in OpenBUGS. The gamma priors with $\xi = 0.01$ and $\xi = 0.001$ did not converge after 1.5 million iterations. For the chains using $\xi = 0.1$ and $\xi = 0.5$, we used 1.5 million including a burn-in of 100,000 iterations in OpenBUGS. Table 5.5 shows the averages of the 100 posterior means, coverages, and 95% credible intervals lengths.

All posterior means are positively biased. As we shall see in Section 5.4, this bias may partially be due to the induced priors on the odds ratios. Of particular concern are the posterior means estimating OR_{CD}. The truth and the means fall on opposite sides of 1 which could yield opposing study conclusions. In the $\xi = 0.1$ case, posterior means for the odds ratios corresponding to relative treatment effects are extremely large. Setting $\xi = 0.5$, versus $\xi = 0.1$, results in posterior means that are much closer than the truth, nearly a

fifth of the extreme mean values resulting from $\xi = 0.1$. However, these posterior results still exhibit large positive bias. Further, the credible interval lengths of all odds ratios are extremely large, resulting in coverages of 1. Clearly, the highly informative gamma priors on τ for modeling between-study heterogeneity needs careful consideration and may not be a good choice for this type of data.

Table 5.4: Posterior results for uniform prior simulations.

True Values		$B = 2$	$B = 10$	$B = 50$	$B = 100$
$OR_{BC} = 1.649$	Mean	1.90	1.97	1.97	1.97
	Median	1.69	1.73	1.74	1.74
	Coverage	0.98	0.97	0.96	0.96
	Length	3.71	3.93	3.81	3.81
$OR_{BD} = 1.350$	Mean	1.88	2.29	2.31	2.31
	Median	1.47	1.67	1.65	1.65
	Coverage	0.98	0.98	0.98	0.98
	Length	5.38	6.64	6.51	6.51
$OR_{CD} = 0.819$	Mean	1.17	1.23	1.17	1.17
	Median	0.90	0.90	0.88	0.88
	Coverage	0.99	0.98	0.98	0.98
	Length	3.37	3.47	3.15	3.15
$OR_{AB} = 1.221$	Mean	1.55	1.39	1.44	1.42
	Median	1.37	1.21	1.26	1.26
	Coverage	0.94	0.98	0.96	0.96
	Length	3.12	2.89	2.86	2.86
$OR_{AC} = 2.014$	Mean	2.38	2.48	2.56	2.56
	Median	2.12	2.17	2.24	2.24
	Coverage	0.99	0.99	0.98	0.98
	Length	4.55	4.98	5.17	5.17
$OR_{AD} = 1.649$	Mean	2.31	3.00	2.99	2.99
	Median	1.81	2.14	2.13	2.13
	Coverage	0.97	0.98	0.98	0.98
	Length	6.42	8.86	8.89	8.89
$\sigma = 0.500$	Mean	0.53	0.57	0.54	0.54
	Median	0.47	0.49	0.46	0.46
	Coverage	0.96	0.96	0.97	0.97
	Length	1.27	1.46	1.41	1.41

5.4 Induced Priors on the Odds Ratios

We have noted the potential danger of the induced prior on the odds ratios. Indeed, the induced prior gives too much weight to extremely large positive values which could skew the posterior in the positive direction. In this section, we discuss the induced priors on the odds ratios and investigate the impact on posterior results in our model (5.3).

Table 5.5: Posterior results for gamma prior simulations.

ξ :		0.001	0.01	0.1	0.5
$OR_{AB} = 1.12$	Mean	NA	NA	3.07	1.77
	Median	NA	NA	1.08	1.13
	Coverage	NA	NA	1.00	1.00
	Length	NA	NA	16.21	6.78
$OR_{AC} = 2.34$	Mean	NA	NA	5.57	3.39
	Median	NA	NA	2.06	2.21
	Coverage	NA	NA	1.00	1.00
	Length	NA	NA	29.44	12.75
$OR_{AD} = 1.49$	Mean	NA	NA	3.65	2.17
	Median	NA	NA	1.31	1.40
	Coverage	NA	NA	1.00	1.00
	Length	NA	NA	19.26	8.23
$OR_{BC} = 2.12$	Mean	NA	NA	42.88	8.90
	Median	NA	NA	2.75	2.71
	Coverage	NA	NA	1.00	1.00
	Length	NA	NA	151.28	41.13
$OR_{BD} = 1.35$	Mean	NA	NA	27.94	6.04
	Median	NA	NA	1.75	1.69
	Coverage	NA	NA	1.00	1.00
	Length	NA	NA	102.66	27.29
$OR_{CD} = 0.64$	Mean	NA	NA	11.92	2.30
	Median	NA	NA	0.79	0.79
	Coverage	NA	NA	1.00	1.00
	Length	NA	NA	38.90	10.39
$\sigma = 0.40$	Mean	NA	NA	2.45	1.35
	Median	NA	NA	2.28	1.25
	Coverage	NA	NA	0.00	0.00
	Length	NA	NA	3.14	1.90

Recall that the mean treatment effects, d_{bk} are log odds ratios so that $OR_{bk} = \exp(d_{bk}) = \exp(d_{Ak} - d_{Ab})$ where d_{Ak} and d_{Ab} are direct effects assumed to be distributed N$(0, \tau_d)$. Therefore, the induced prior on the odds ratio corresponding to direct effects, OR_{Au}, is log-normal with mean 0 and precision τ_d. The mean relative effects, d_{bk}, are N$(0, \tau_d/2)$. Thus, the induced prior on the odds ratio is $OR_{bk} \sim$ LN$(0, \tau_d/2)$. When $\tau_d = 0.1$, as in Section 5.2.2, this induced prior is nearly uniform over 0 to 50,000. Setting $\tau_d = 0.0001$, as in Section 5.2.1, induces a prior on OR_{bk} that is approximately uniform from 0 to 10^{100}.

At first thought, the uniform nature of the induced prior may seem appealing because of the uniform's typical use as a non-informative prior. However, this prior, LN$(0, \tau_d/2)$, gives as much probability to extremely large values as to decimal values of OR_{bk}. Thus, OR_{bk} is equally likely to take on values of, say, $45,000$ as it is to be 0.01. An odds ratio of $45,000$ is impractical, to say the least.

Further, the induced prior, say LN$(0, 0.0001/2)$, gives much greater probability to values of OR_{bk} that are greater than 1 than below 1. This is problematic considering a typical study success criterion depends on which side of 1 OR_{bk} is. To see this simply, we approximate LN$(0, 0.0001/2)$ as a uniform prior on 0 to 10^{100}. Then, the prior probability that $OR_{bk} < 1$ is 10^{-100}. The prior probability that $OR_{bk} > 1$ is equal to $1 - 10^{-100}$ which is so large that it is approximately 1. Figure 5.7 illistrates this graphically. The graph of the induced prior on OR_{bk} is amplified, displayed from 0 to 100, in order to see the small shaded probability that OR_{bk} is less than 1.

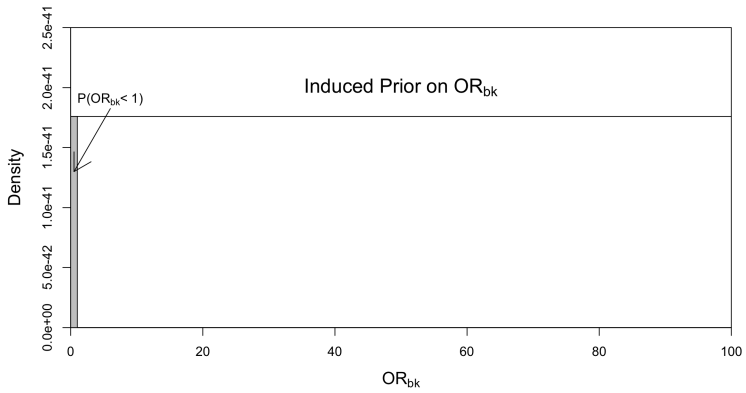


Figure 5.7: The induced prior, displayed from 0 to 100, on the odds ratio. The shaded region denotes the probability that the odds ratio is less than 1.

5.4.1 Effect on Posterior Results

In this section, we use simulated data sets to investigate the consequences of the odds ratio's induced prior on the posterior results. The simulated data follows the model in (5.3) with $\tau_\mu = 0.1$ and $N_S = 9$ studies, each with a sample size between 50 and 100. The sample size for each study was chosen by generating a random number from U$(50, 100)$. The mean μ was generated from a N$(-3, 0.25)$ distribution, and we set $d_{AB} = 0.2$, $d_{AC} = 0.7$, and $d_{AD} = 0.5$. In effect, these simulated data are similar to those resulting in Table 5.2 in Section 5.2.2.

Recall in previous section we used normal priors on the direct treatment effects d_{Ad} and d_{Ak} with mean 0 and precision, τ_d, equal to 0.0001 or 0.1. As discussed above, these result in unrealistically diffuse induced priors on the odds ratios. Here, we consider normal priors with 0 means and precisions, τ_d, of 0.25, 1, and 2. The priors are shown in Figure 5.8. The induced prior on the odds ratio resulting from the 0.25 precision is extremely diffuse, with its right tail stretching to near one million.

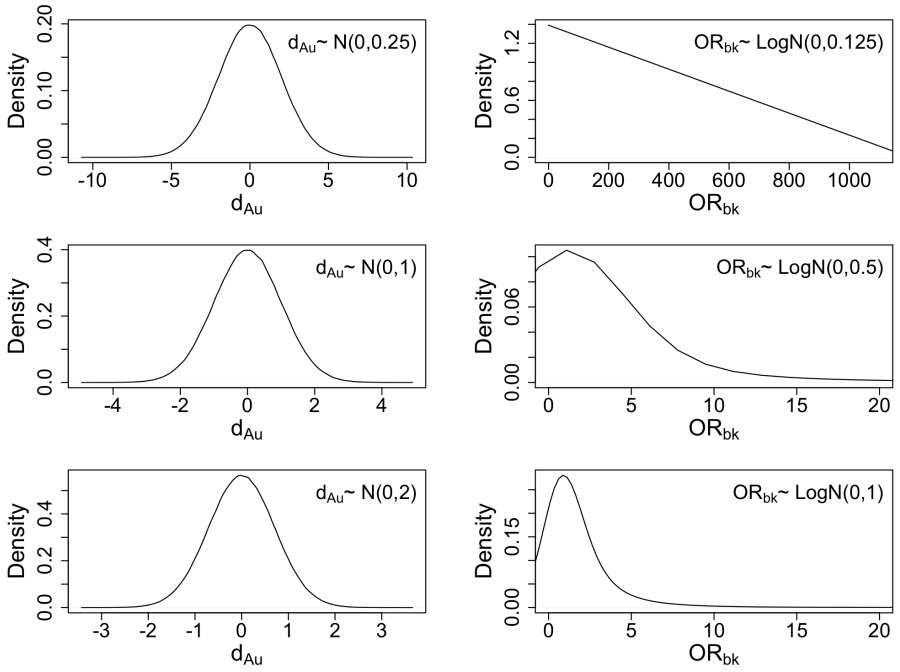


Figure 5.8: Induced priors on the odds ratio given diffuse normals on the treatment effects.

We ran the simulations in JAGS using 3.5 million iterations including 250,000 burn-in and thinned every 50. The only exception was the gamma case with the N(0, 0.25) prior, which ran in OpenBUGS with 500,000 iterations including a burn-in of 100,000 and thin of 10. The average of the 100 posterior means, medians, 95% credible interval lengths, and coverages are in Table 5.6. The gamma case with the N(0, 0.25) prior behaves the worst with extremely large posterior means and credible interval lengths. As expected, the posterior means are positively skewed. Indeed, the posterior medians, for this case, all under-estimate the true odds ratios.

Of particular interest are the results for OR_{CD}. This odds ratio is below one. It is important for the analysis to capture that because, often, a study's success depends on which side of one an odds ratio falls. As the prior on the treatment effects gets less diffuse, the posterior means of OR_{CD} get closer to 1 but still do not fall below 1, as we would hope. The posterior medians are better estimates in most cases. Indeed the posterior medians of OR_{CD} are below 1 in most cases, though not far below 1.

The posteriors for the odds ratios in the N(0, 0.25) case were skewed by such large outliers, that the densities were unrecognizable. In all odds ratios, the tails of the corresponding posterior distributions were extremely stretched in the positive direction, so much so that the peak of the density was squeezed into the far left and not interpretable. Figure 5.9 displays posterior densities on OR_{BD} in the first simulated data set. Normal priors (in terms of mean and precision) on corresponding the treatment effect $d_{BD} = \log(OR_{BD})$ are noted in top right corners. From left to right, normal priors on d_{BD} are increasingly diffuse making the posteriors on OR_{BD} increasingly skewed right.

Table 5.6: Posterior simulation results regarding the induced prior on the odds ratio.

		$d_{Au} \sim$ N$(0, 0.25)$			$d_{Au} \sim$ N$(0, 1)$			$d_{Au} \sim$ N$(0, 2)$		
Prior on heterogeneity:		U	G	LN	U	G	LN	U	G	LN
$OR_{AB} = 1.221$	Mean	1.68	6.65	1.63	1.41	1.63	1.41	1.29	1.28	1.29
	Median	1.31	0.99	1.31	1.22	0.99	1.22	1.16	1.00	1.15
	Coverage	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	Length	4.86	45.28	4.43	3.17	6.82	3.15	2.46	3.72	2.48
$OR_{AC} = 2.014$	Mean	2.82	6.95	2.77	2.14	1.65	2.14	1.80	1.29	1.79
	Median	2.24	1.03	2.25	1.88	1.01	1.86	1.64	1.00	1.61
	Coverage	0.98	1.00	0.98	0.99	1.00	1.00	0.97	1.00	0.97
	Length	7.84	47.26	7.35	4.59	6.93	4.62	3.33	3.75	3.36
$OR_{AD} = 1.649$	Mean	4.06	6.64	3.77	2.42	2.73	2.43	1.91	1.66	1.91
	Median	2.31	0.99	2.33	1.81	1.02	1.81	1.55	1.01	1.53
	Coverage	0.98	1.00	0.98	1.00	1.00	0.99	1.00	1.00	1.00
	Length	17.11	45.18	15.19	7.78	15.90	7.80	5.19	6.97	5.26
$OR_{BC} = 1.649$	Mean	2.19	48.81	2.16	1.77	1.63	1.77	1.55	1.28	1.54
	Median	1.75	1.05	1.76	1.55	1.00	1.54	1.40	1.00	1.39
	Coverage	0.98	1.00	0.99	0.98	1.00	0.99	0.98	1.00	0.99
	Length	6.09	235.39	5.73	3.87	6.85	3.90	2.92	3.72	2.95
$OR_{BD} = 1.350$	Mean	3.06	45.64	2.83	2.00	2.69	2.00	1.65	1.65	1.65
	Median	1.76	1.00	1.78	1.48	1.00	1.48	1.32	1.00	1.32
	Coverage	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Length	12.64	224.89	11.32	6.47	15.71	6.50	4.53	6.93	4.59
$OR_{CD} = 0.819$	Mean	1.75	43.46	1.59	1.32	2.65	1.31	1.19	1.63	1.19
	Median	1.03	0.96	1.03	0.98	0.99	0.98	0.96	0.99	0.96
	Coverage	0.98	1.00	0.98	1.00	1.00	0.99	1.00	1.00	1.00
	Length	7.04	213.89	6.08	4.28	15.44	4.20	3.28	6.87	3.29
$\sigma = 0.50$	Mean	0.85	16.95	0.82	0.78	16.97	0.79	0.78	16.96	0.79
	Median	0.75	16.05	0.77	0.69	16.06	0.74	0.69	16.04	0.74
	Coverage	0.96	0.00	0.97	0.98	0.00	0.97	0.98	0.00	0.98
	Length	2.12	18.14	1.38	1.92	18.28	1.31	1.88	18.26	1.29

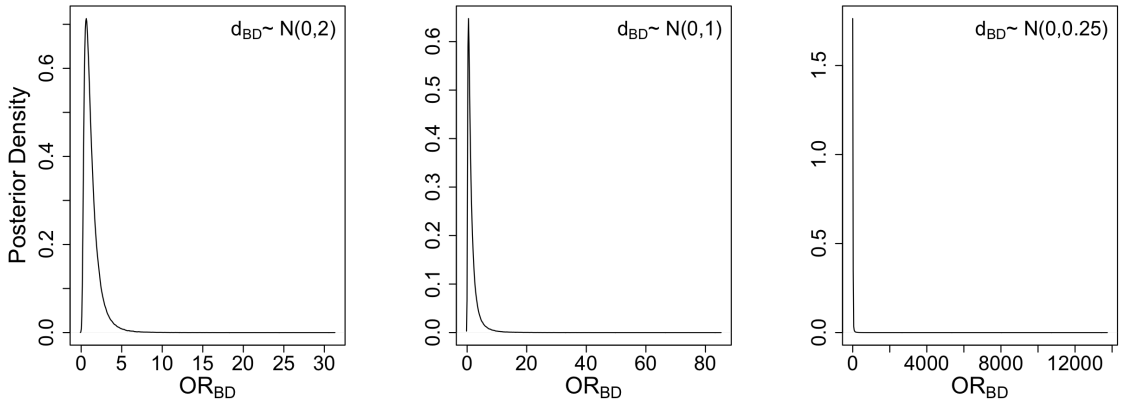


Figure 5.9: Posterior densities for OR_{BD}. Normal priors (in terms of mean and precision) on corresponding the treatment effect $d_{BD} = \log(OR_{BD})$ are shown in top right corners.

5.5 Meta-Regression to Model the Baseline

In this section, we extend the model in (5.3) to include a covariate C so that:

$$\text{logit}(p_{jk}) = \underline{\mu_{jb} + \beta C_j} + \delta_{jbk}\text{I}\{k \text{ after } b\} \tag{5.5}$$

where $\text{I}\{k \text{ after } b\}$ equals 1 if the kth study is "after" the bth and 0 otherwise. The underlined section of (5.3) is considered the baseline model. The baseline model is also known as the underlying risk and may explain study characteristics such as patient age and medical history. Such underlying characteristics can affect treatment. While random effects models can be used to capture such between-trial heterogeneity, the inclusion of covariates help to explain the source of the heterogeneity. This is known as meta-regression. In this section, we investigate the use of meta-regression on the posterior results. We run simulated data sets using the full model in (5.5), a model ignoring the covariate, and a model ignoring the baseline effects. For the model ignoring the covariate, we eliminate βC_j from (5.5). To ignore the baseline effects, we give each μ_{jb} an independent diffuse normal prior with mean 0 and precision 0.1.

5.5.1 Meta-Regression Simulations

We simulate 100 datasets each with a sample size of 100 corresponding to the model in (5.5) with $\beta = 3$. As before, we consider 4 treatments, but we increase the number of studies to $N_S = 30$. The true value of μ is simulated from a N$(-1, 0.25^2)$ distribution. Also, we deviate slightly from the model in (5.3) by giving μ a heirarchical model. Specifically, $\mu \sim N(m, \tau_m)$ where $\tau_m = 1/\sigma_m$. The hyperparameters m and σ_m are given N$(0, 10^2)$ and U(0,2) priors, respectively. We fit the baseline and relative treatment effects simultaneously using in WinBUGS. The WinBUGS models are in the appendix. Table 5.7 displays the average of 100 posterior means, coverages, and 95% credible interval lengths using 4 million MCMC iterations in JAGS with a burn-in of 200 thousand and thinning every 50 to alleviate autocorrelation.

Posterior means for all of the odds ratios, except for OR_{AB}, are closest to the truth when using the full model. The unexpected result for OR_{AB} is most likely due to MCMC error and suggests that there is no real difference between the models for for OR_{AB}. Overall, the model ignoring the covariate performs slightly better than the one ignoring the baseline in terms of posterior means. We may begin to see more divergence between these results with more simulations with larger sample sizes.

Posterior credible interval lengths widen from the full model to the model ignoring the covariate to the model ingoring the baseline. This is expected because there is increased uncertainty in each model. The average coverage is high, above 95%, in each model for all parameters. Also, the median is closer to the truth than the mean in all cases.

Table 5.7: Posterior results for $n = 100$, $N_S = 30$, and $\beta = 3$.

True Values		Full model	Ignoring covariate	Ignoring baseline
$OR_{AB} = 1.22$	Mean	1.374	1.371	1.369
	Median	1.277	1.271	1.269
	Coverage	0.970	0.990	0.970
	Length	2.054	2.092	2.105
$OR_{AC} = 2.01$	Mean	2.276	2.298	2.299
	Median	2.106	2.123	2.124
	Coverage	0.960	0.970	0.960
	Length	3.512	3.582	3.600
$OR_{AD} = 1.65$	Mean	2.085	2.150	2.166
	Median	1.797	1.843	1.853
	Coverage	0.960	0.950	0.960
	Length	4.541	4.777	4.853
$OR_{BC} = 1.65$	Mean	1.811	1.845	1.845
	Median	1.676	1.702	1.703
	Coverage	0.960	0.960	0.960
	Length	2.767	2.879	2.885
$OR_{BD} = 1.35$	Mean	1.688	1.756	1.767
	Median	1.455	1.501	1.509
	Coverage	0.960	0.960	0.960
	Length	3.673	3.933	3.980
$OR_{CD} = 0.82$	Mean	1.027	1.063	1.068
	Median	0.884	0.905	0.908
	Coverage	0.950	0.950	0.940
	Length	2.233	2.412	2.438
$\sigma = 0.500$	Mean	0.569	0.561	0.564
	Median	1.277	1.271	1.269
	Coverage	0.950	0.970	0.970
	Length	0.931	0.974	0.980

5.6 Conclusion

In this chapter, we used simulated logistic datasets to investigate the affect of prior choices on the posterior results. We focused on prior specifications regarding the between-trial heterogeneity. We looked at four commonly-used diffuse priors in Section 5.2 and concentrated on parameter choices for the uniform and gamma priors in Section 5.3. In Section 5.4, we found that the induced prior on the odds ratio may cause positive bias. In Section 5.5, we considered meta-regression and consequences of ignoring the covariate and/or the baseline in the baseline model.

Overall, the resulting posterior means were substantially biased in the positive direction for all simulations with little data. To simulate datasets with little data, we used logistic data with small means and relatively few studies with small sample sizes. Further credible interval lengths were large overall. Our simulations provide evidence that much care should be taken in modeling small logistic datasets with few successes in meta-analyses. In particular, the extremely informative prior on τ and the diffuse normal priors on the treatment effects are problematic. However, there were no priors that performed well for these data. Perhaps even fixed-effects models would be a wise choice in cases such as these.

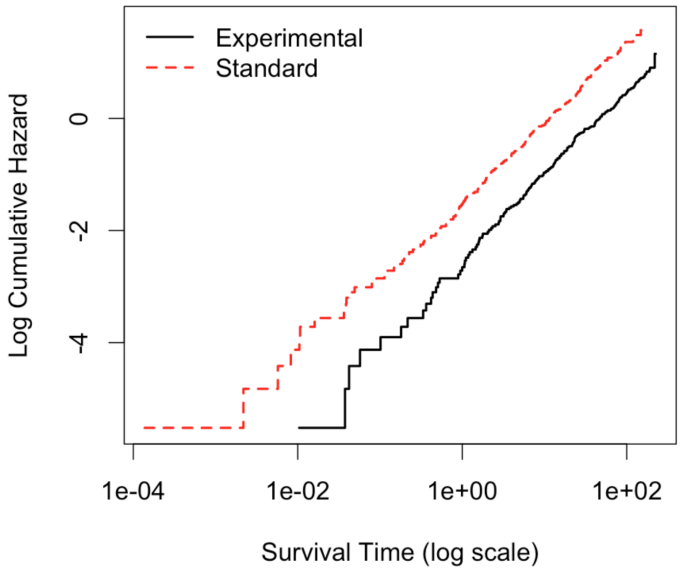
APPENDICES

APPENDIX A

Chapter Two Appendix

A.1 Parametric PHM Assumption Checks

As discussed in Klein and Moeschoberger [33], the proportional hazards and the parameteric survival times assumption can be tested graphically. Throughout the examples in this chapter, we used an `R` function called `WeibullDiag` to do so. Note that the exponential distribution is a specific case of the Weibull distribution, and the Weibull PHM can be expressed as an accelerated failure time (AFT) model. Details on this are discussed in Section 3.1. The AFT model represents the log cumulative hazard as a linear function of the log survival times. Therefore, to test the Weibull (or exponential) PHM assumptions, the log cumulative hazard (produced using Kaplan-Meier estimates) is plotted against the log survival times for each representation of the categorical variable. If the assumptions hold, the plot should be linear and parallel. An example diagnostic plot is shown below. The example plot shows that the Weibull PHM assumptions hold.



```
# input expert best guesses in ascending order
# For one binary covariate model, only 2 modes
modes <- c(65, 60, 45, 40)

# input expert upper bounds in ascending order
# For one binary covariate model, only 2 pctiles
pctiles <- c(92, 87, 75, 70)

# input the percentile to which to set the upper bounds
pct <- .75

# find gamma parameters for priors on conditional means (m)
# or medians (tm) using the expert info
xi1 <- xi2 <- NULL
for(i in 1:length(modes)){
  xi1[i] <-  elicit_gamma(modes[i], pct, pctiles[i],
  "less" )[1]
  xi2[i] <-  elicit_gamma(modes[i], pct, pctiles[i],
  "less" )[2]
}
xi1; xi2

# sample from the gamma priors
# on condtional means (m) or medians (tm)
n <- 100000
m <- matrix(NA, n, length(modes))
```

```
tm <- matrix(NA, n, length(modes))
for(i in 1:length(modes)){
  m[,i] <- rgamma(n, shape = xi1[i],
  scale = xi2[i]) # means
  tm[,i] <- rgamma(n, shape = xi1[i],
  scale = xi2[i]) # medians
}

##### EXP CMP ######
# sample induced values
e.m.beta <- matrix(NA, n, length(modes))
e.m.HR <- matrix(NA, n, length(modes))
e.m.gam <- 1/m[,1]
for(i in 2:length(modes)){
  e.m.beta[,i] <- log(1/(e.m.gam*m[,i]))
  e.m.HR[,i] <- exp(e.m.beta[,i])}

##### EXP MEDIAN ######
# sample induced values
e.med.beta <- matrix(NA, n, length(modes))
e.med.HR <- matrix(NA, n, length(modes))
e.med.gam <- log(2)/tm[,1]
for(i in 2:length(modes)){
  e.med.beta[,i] <- log(log(2)/(e.med.gam*tm[,i]))
  e.med.HR[,i] <- exp(e.med.beta[,i])}

###### WEIB CMP #########
```

```
# sample induced values
k=3; lambda<-runif(n, 0, (log(2)/k))
w.m.beta <- matrix(NA, n, length(modes))
w.m.HR <- matrix(NA, n, length(modes))
w.m.r=c(); m0 <- m[,1]
for(i in 1:length(m0)){
  f<- function(x)
  {-((x^(-1))*log(lambda[i]))+lgamma(1+(1/x))-log(m0[i])}
  w.m.r[i]= uniroot.all(f, c(0.01,100000))}
for(i in 2:length(modes)){
  w.m.beta[,i]=log(((m[,i]/gamma(1+(1/w.m.r)))^
  (-w.m.r))/lambda)
  w.m.HR[,i]<-exp(w.m.beta[,i])}

###### WEIB MEDIANS ##########
# sample induced values
k=3; lambda<-runif(n, 0, (log(2)/k))
w.med.beta <- matrix(NA, n, length(modes))
w.med.HR <- matrix(NA, n, length(modes))
for(i in 2:length(modes)){
  w.med.r <-(log(log(2)/lambda))/(log(tm[,1]))
  w.med.beta[,i] <- log((log(2))/(lambda*
  ((tm[,i])^w.med.r)))
  w.med.HR<-exp(w.med.beta[,i])}

\section{OpenBUGS Models for Posterior Analysis}
###  EXPONENTIAL PHM WITH ONE BINARY COVARIATE ###
```

```
# OpenBUGS model
ephm <- function(){
    tm0 ~ dgamma(k0, s0)
    tm1 ~ dgamma(k1, s1)
    s0 <- 1/theta0
    s1 <- 1/theta1
    gamma <- log(2)/tm0
    beta <- log((log(2))/(gamma*tm1))

# Approximate Independent Priors
#  gamma ~ dlnorm(-4.65, 9.77)
 # beta ~ dnorm(-.127, 3.97)

  # Likelihood of the survival time data
  for (j in 1:N) {
    HR[j] <- exp(Z[j]*beta)
    rate[j] <- gamma*HR[j]
    t[j] ~ dexp(rate[j])%_%I(t.cen[j],)
    # the "%_%" is a dummy to prevent R errors
  }
}

##### WEIB PHM WITH ONE BINARY COVARIATE ####
# OpenBUGS model
wphm <- function(){
# priors
 tm0 ~ dgamma(k0, s0)
```

```
 tm1 ~ dgamma(k1, s1)
 s0 <- 1/theta0
 s1 <- 1/theta1
  lambda ~ dunif(0, B)
  B <- log(2)/d

# induced priors
#  r <- (log(log(2)/lambda))/(log(tm0))
#  beta <- log(log(2)/(lambda*pow(tm1,r)))

# Approximate Independent Priors
#  r ~ dexp(4.31)%_%T(.3,) # "%_%" is a dummy
# beta ~ dnorm(-0.07, 22.7) # precision

  # Likelihood of the survival time data
  for (j in 1:N) {
    HR[j] <- exp(Z[j]*beta)
    Scale[j] <- lambda*HR[j]
    t[j] ~ dweib(r,Scale[j])%_%I(t.cen[j],)
  }
}
```

APPENDIX B

Chapter Three Appendix

B.1 Example of Prior ESS Calculation Using Poisson Sampling

Morita et al. define the prior ESS as the m that minimizes the distance

$$|D_p(\lambda) - D_q(m, \lambda, y_m)| \tag{B.1}$$

where$D_p(\lambda)$ and $D_q(m, \lambda, y_m)$ are defined as follows.

$$D_p(\lambda) = \frac{-\partial}{\partial \lambda^2} \log(\pi(\lambda \mid \tilde{\lambda})) \tag{B.2}$$

where $\pi(\lambda \mid \tilde{\lambda})$ is the prior on the parameter of interest λ given hyperparameters $\tilde{\lambda}$.

$$D_q(m, \lambda, y_m) = \frac{-\partial}{\partial \lambda^2} \log(q_m(\lambda \mid \lambda_0, y_m)) \tag{B.3}$$

where $q_m(\lambda \mid \lambda_0, y_m) \propto q_0(\lambda \mid \tilde{\lambda}) \prod i = 1^m f(y \mid \lambda)$ given the hypothetical vague prior $q_0(\lambda \mid \tilde{\lambda})$ (Morita et al. provide suggestions for the hypothetical vague prior) and $f(y \mid \lambda)$ the likelihood of the current data.

The hypothetical prior follows the suggestion in Morita et al.:

$$q_0(\lambda \mid \tilde{\lambda}) = \lambda^{a_0 \sum y_{0i}/c} \exp(\frac{-a_0 n_0 \lambda}{c}) \tag{B.4}$$

where c is large in order to inflate the variance. That is, the vague prior is a gamma with the same hyperparameters as the power prior divided by a constant c. Therefore,

$$D_p(\lambda) = \frac{a_0 \sum y_{0i}}{\lambda^2} \tag{B.5}$$

and

$$\begin{aligned} D_q(m, \lambda, y_m) &= \frac{-\partial}{\partial \lambda^2} \log \left\{ \lambda^{a_0 \sum y_{0i}/c} \exp(\frac{-a_0 n_0 \lambda}{c}) \exp(-\lambda m) \lambda^{\sum_{i=1}^{m} y_i} \right\} \\ &= \frac{a_0 \sum y_{0i}}{c\lambda^2} - \frac{\sum_{i=1}^{m} y_i}{\lambda^2} \end{aligned} \tag{B.6}$$

So, the distance $\delta(m)$ to be minimized for the prior ESS calculation is

$$\delta(m) = \left| \frac{1}{c\lambda^2} \left((c-1)\, a_0 \sum_{i=1}^{n_0} y_{0_i} - c \sum_{i=1}^{m} y_i \right) \right| \tag{B.7}$$

APPENDIX C

Chapter Four Appendix

C.1 Sensitivity Indicatrices for Lower Percentiles

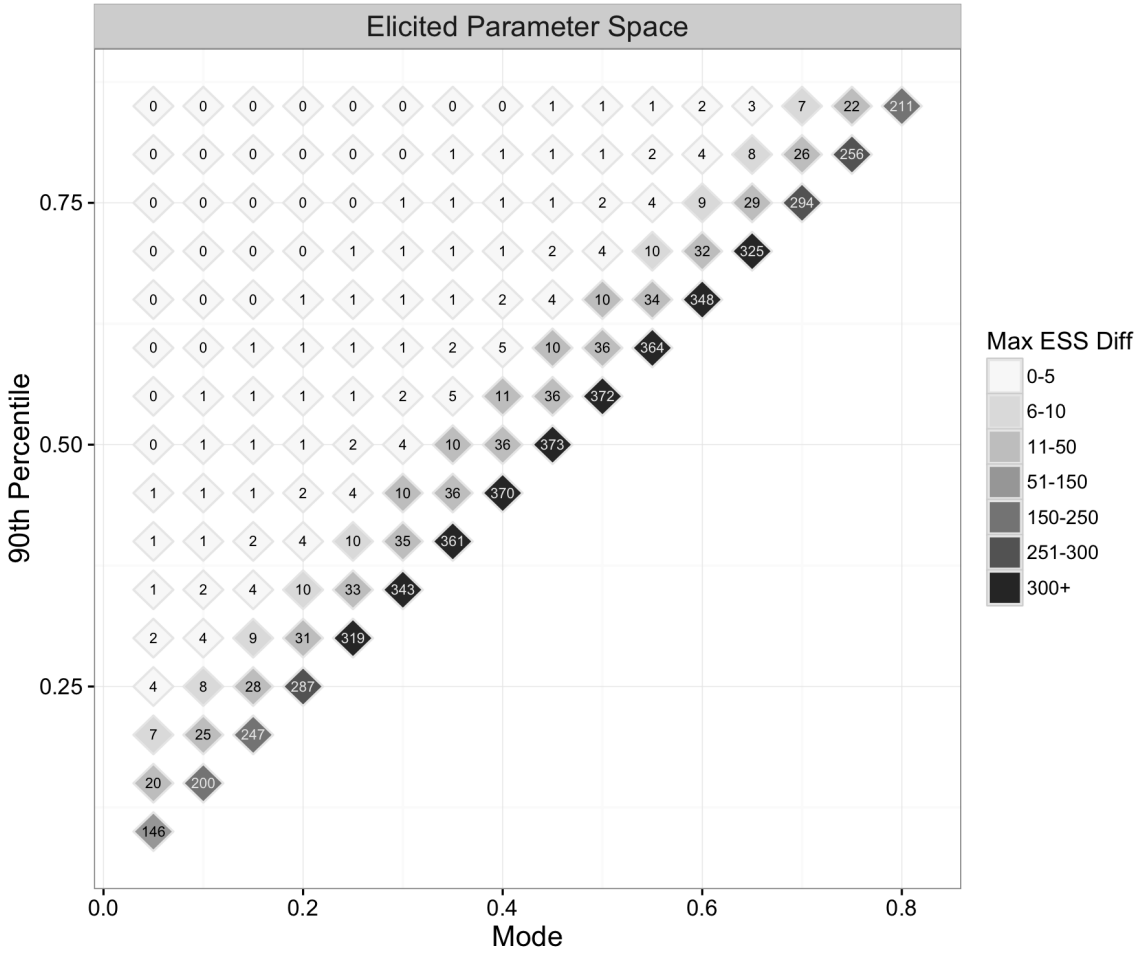


Figure C.1: Sensitivity indicatrix using the ℓ_1 metric and $\xi = 0.90$.

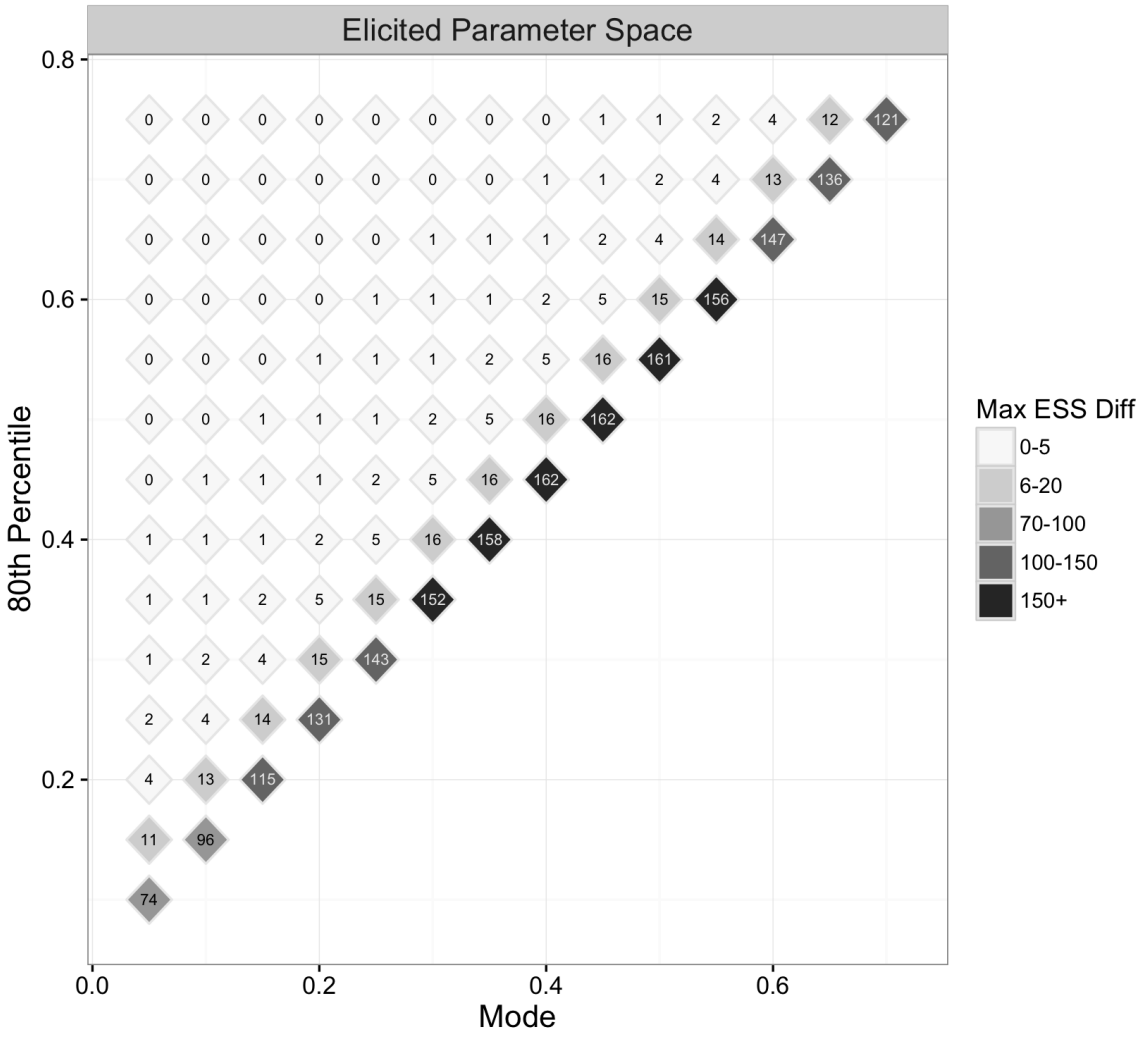


Figure C.2: Sensitivity indicatrix using the ℓ_1 metric and $\xi = 0.80$.

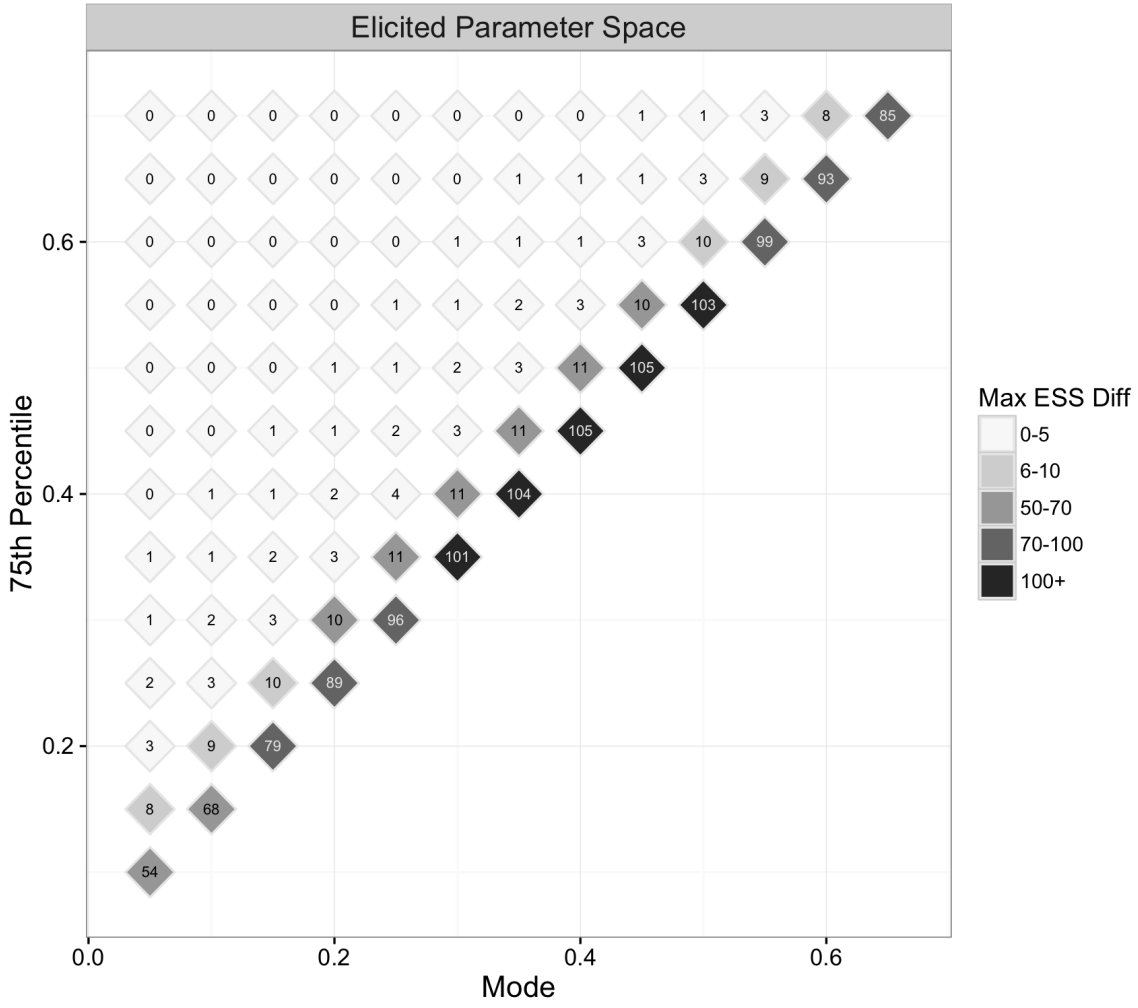


Figure C.3: Sensitivity indicatrix using the ℓ_1 metric and $\xi = 0.75$.

C.2 Code – Elicitation Maps

```
library(glmcmp)

library(ggrepel)

library(dplyr)

library(grid)


n <- 100

r <- .02

pct <- .95; mm <- .9
```

```
make_l2_ball <- function(center, radius=r, n=n){
  x <- center[1]; y <- center[2]
  t <- seq(0, 2*pi, length.out = n+1)[-(n+1)]
  data.frame(x = x + radius*cos(t), y = y + radius*sin(t))
}

##### make mode-percentile grid, look at image
centers <- expand.grid(
  mode = seq(.05, mm, .05),
  pctile = seq(.05, mm, .05)
) %>%
  filter(mode < pctile) %>% filter(round(mode,2) != round(pctile,2))
# any mode is less than or equal to its corresponding (pctile - 0.5)

labels <- centers$label <- 1:nrow(centers)

centers_image <- centers %>%
  select(mode, pctile) %>%
  apply(1, function(v){
    v <- unname(v)
    elicit_beta(mode = v[1], pctile = v[2], pct = pct,
                side = "less", U = 1000)
  }) %>% t %>% as.data.frame %>%
  mutate(label = labels)

centers_image$label <- labels

centers_df <- data.frame(mode_cent = centers$mode,
pctile_cent=centers$pctile,
```

```
centerESS= centers_image$alpha+centers_image$beta)



l2_balls <- centers %>%
  select(mode, pctile) %>%
  apply(1, make_l2_ball, radius = r, n = n) %>%
  bind_rows %>%
  mutate(label = rep(labels, each = n))

# transform all of the circle boundary points into the
# standard parameter space (alpha-beta)
l2_balls_images <- l2_balls %>%
  select(-label) %>%
  apply(1, function(v){
    v <- unname(v)
    elicit_beta(mode = v[1], pctile = v[2], pct = pct,
                side = "less", U = 1000)
  }) %>% t %>% as.data.frame %>%
  mutate(label = rep(labels, each = n))



#### Calculate ESSs
images_df <- mutate(l2_balls_images,
"ESS"=l2_balls_images$alpha + l2_balls_images$beta)
maxESSdiff <- NULL
for(i in 1:nrow(centers)){
  images <- filter(images_df, images_df$label==i)
  center <- filter(centers_image, centers_image$label==i)
```

```
  centerESS <- center$alpha + center$beta
  maxESSdiff[i] <- max(abs(images$ESS-centerESS))
}
```

BIBLIOGRAPHY

[1] A. E. Ades, D. Caldwell, S. Reken, N. Welton, A. Sutton, and S. Dias. Evidence synthesis for decision making 7: A reviewer's checklist. *Medical Decision Making*, 33(5):679–691, 2013.

[2] US Food & Drug Administration. Guidance for the use of bayesian statistics in medical device clinical trials. http://www.fda.gov/RegulatoryInformation/Guidances/ucm071072.htm, 2010.

[3] Jim Albert. *Bayesian Computation with R*. Springer, New York, 2009.

[4] E. Bedrick, R. Christensen, and W. Johnson. A new perspective on priors for generalized linear models. *Journal of the American Statistical Association*, 91(436):1450–1460, 1996.

[5] Canavos and Tsokos. Bayesian estimation of life parameters in the weibull distribution. *Journal of Operations Research*, 21(3):755–763, 1973.

[6] John B. Carlin. Meta-analysis for 2x2 tables: A bayesian approach. *Statistics in Medicine*, 11:141–158, 1992.

[7] K. Chaloner, T. Church, T. Louis, and J. Matts. Graphical elicitation of a prior distribution for a clinical trial. *The Statistician*, 42:341–353, 1993.

[8] M. Chen and J. Ibrahim. The relationship between the power prior and hierarchical models. *Bayesian Analysis*, 1(3):551–574, 2006.

[9] M. Chen, A. Manatunga, and C. Williams. Heritability estimates from human twin data by incorporating historical prior information. *Biometrics*, 54:1348–1362, 1998.

[10] Christensen, Johnson, Branscum, and Handson. *Bayesian Ideas and Data Analysis*. Taylor and Francis Group, Boca Raton, FL, 2011.

[11] R.M. Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, New York, 1991.

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

[13] S. Dias, A. Sutton, A. E. Ades, and N. Welton. Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Medical Decision Making*, 33(5):607–617, 2013.

[14] S. Dias, A. Sutton, N. Welton, and A. E. Ades. Evidence synthesis for decision making 3: Heterogeneity- subgroups, meta-regression, bias, and bias-adjustment. *Medical Decision Making*, 33(5):618–640, 2013.

[15] S. Dias, N. Welton, A. Sutton, and A. E. Ades. Evidence synthesis for decision making 1: Introduction. *Medical Decision Making*, 33(5):597–606, 2013.

[16] S. Dias, N. Welton, A. Sutton, and A. E. Ades. Evidence synthesis for decision making 5: The baseline natural history model. *Medical Decision Making*, 33(5):657–670, 2013.

[17] S. Dias, N. Welton, A. Sutton, and A. E. Ades. Evidence synthesis for decision making 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis. *Medical Decision Making*, 33(5):671–678, 2013.

[18] S. Dias, N. Welton, A. Sutton, D. Caldwell, L. Guobing, and A. E. Ades. Evidence synthesis for decision making 4: Inconsistency in networks of evidence based on randomized controlled trials. *Medical Decision Making*, 33(5):641–656, 2013.

[19] Y. Duan, E. Smith, and K. Ye. Evaluating water quality using power priors to incorporate historical information. *Environmetrics*, 17:95–106, 2006.

[20] Y. Duan, E. Smith, and K. Ye. Using power priors to improve the binomial test of water quality. *Journal of Agricultural, Biological, and Environmental Stastistics*, 11:151–168, 2006.

[21] Y. Duan and K. Ye. Normalized power prior bayesian analysis. Working paper series, The U. of Texas at San Antonio, College of Business, Wp#0058MSS-432-2008., 2008.

[22] Pasquale Erto and Giorgio Massimiliano. A note on using bayes priors for weibull distribution. arXiv:1310.7056 [math.ST] Cornell U. Library (as of 26 October 2013), 2011.

[23] Paul H. Garthwaite, Joseph B. Kadane, and Anthony O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100:470:680–701, 2005.

[24] Andrew Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.

[25] Michael S. Hamanda, Alyson Wilson, C. Shane Reese, and Harry Martz. *Bayesian Reliability*. Springer, New York, 2008.

[26] J. Ibrahim and M. Chen. Power prior distributions for regression models. *Statistical Science*, 15:46–60, 2000.

[27] J. Ibrahim, M. Chen, and Sinha D. *Bayesian Survival Analysis*. Springer-Verlag, New York, 2001.

[28] J. Ibrahim, M. Chen, and L. Ryan. Bayesian variable selection for time series count data. *Statistica Sinica*, 10:971–987, 2000.

[29] John W. Seaman III, John W. Seaman Jr., and James D. Stamey. Hidden dangers of specifying noninformative priors. *The American Statistician*, 66(2):77–84, 2012.

[30] J. B. Kadane, J. Dickey, R. Winkler, W. Smith, and S. Peters. Interactive elicitation of opinion for a normal linear model. *Journal of the American Statistical Association*, 75:845–854, 1980.

[31] J. B. Kadane and L. J. Wolfson. Experiences in elicitation. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47:3–19, April 1998.

[32] R. Kass and L. Wasserman. The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(435), 1996.

[33] Klein and Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, 2003.

[34] A Legedza and J. Ibrahim. Heterogeneity in phase i clinical trials: prior elicitation and computation using the continual reassessment method. *Statistics in Medicine*, 20, 2001.

[35] S. Low-Choy, James A. Murray, and K. Mengersen. Indirect elicitation from ecological experts: from methods and software to habitat modelling and rock-wallabies. In A. O'Hagan and M. West, editors, *Handbook of Applied Bayesian Analysis*, pages 511–544. Oxford University Press, New York, 2010.

[36] S. Morita, P. Thall, and P. Muller. Determining the effective sample size of a parametric prior. *Biometrics*, 64(2):595–602, 2008.

[37] Satoshi Morita, Peter F. Thall, and Peter Mueller. Evaluating the impact of prior assumptions in bayesian biostatistics. *Statistics in Biosciences*, 2(1):1–17, 2010.

[38] National Institute for Health and Clinical Evidence. Evidence synthesis tsd series. http: //www.nicedsu.org.uk/Evidence-Synthesis-TSD-series

[39] B. Neelon and A. O'Malley. Bayesian analysis using power priors with application to pediatric quality of care. *J. of Biometrics and Biostatistics*, 1:103, 2010.

[40] B. Neuenschwander, M. Branson, and D. Spiegelhalter. A note on the power prior. *Statistics in Medicine*, 28:3562–3566, 2009.

[41] A. et al. O'Hagan. *Uncertain Judgements: Eliciting Experts' Probabilities*. Wiley, New York, 2006.

[42] Ingram Olkin and Ruixue Liu. A bivariate beta distribution. *Statistics & Porbability Letters*, 62:407–412, 2003.

[43] G. Parmigiani. *Modeling in Medical Decision Making: A Bayesian Approach*. Wiley, New York, 2002.

[44] C. Rietbergen, R. Groenwold, H. Hoijtink, K. Moons, and I. Klugkist. Expert elicitation of study weights for bayesian analysis and meta-analysis. *Journal of Mixed Methods Research*, 2014.

[45] C. Rietbergen, I. Klugkist, K. Janssen, K. Moons, and J. Hoijtink. Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemporary Clinical Trials*, 32:848–855, 2011.

[46] H Rinne. *The Weibull Distribution*. Taylor and Francis Group, Boca Raton, FL, 2009.

[47] C. Robert. *The Bayesian Choice*. Springer Verlag, New York, 2nd edition, 2001.

[48] R. Soland. Bayesian analysis of the weibull process with unknown scale and shape parameters. *Advanced Research Department Technical Paper RAC-TP-359*, 1969.

[49] D. Speigelhalter, K. Abrams, and J. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, New York, 2004.

[50] David Spiegelhalter, Andrew Thomas, Nicky Best, and Dave Lunn. Winbugs user manual. http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf, 2003.

[51] D. Stangl and D. Berry. Meta-analysis: Past and present challenges. In D. Stangl and D. Berry, editors, *Meta-Analysis in Medicine and Health Policy*, pages 1–28, New York, 2004. Dekker.

[52] N. Welton, A Sutton, N Cooper, K Abrams, and A Ades. *Evidence Synthesis for Decision Making*. Wiley, New York, 2012.

[53] Y. Wu, W. Shih, and D. Moore. Elicitation of a beta prior for bayesian inference in clinical trials. *Biometrical Journal*, 2:212–223, 2008.