Bayesian methods for the evaluation of Tritium: Relative Biological Effectiveness and cancer risk

Ghassan B. Hamra

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology

Chapel Hill
2011

Approved by:

Dr. David Richardson, Advisor

Dr. Richard Maclehose, Reader

Dr. Steve Wing, Reader

Dr. Robert Millikan, Reader

Dr. John Dement, Reader

ABSTRACT

Ghassan Badri Hamra: Bayesian methods for the evaluation of Tritium: Relative Biological Effectiveness and cancer risk.

(Under the direction of Dr. David Richardson)

Tritium is a radioisotope of hydrogen that releases β^- energy, which is a form of ionizing radiation. Tritium is understudied due to a lack of epidemiological data on human exposure, despite the fact that exposure to low energy ionizing radiation is ubiquitous in the environment. In occupational studies of nuclear workers, tritium is usually aggregated with gamma radiation and examined under the assumption that the cancer risk per unit exposure of each is equivalent. However, a recent systematic review of the literature suggests that beta radiation is more biologically effective than gamma radiation.

We utilize Bayesian methods to inform estimation of cancer risk for tritium as well as the RBE of tritium compared to gamma radiation using information compiled at the Savannah River Site (SRS) nuclear fuel facility in Aiken, SC. SRS staff maintained detailed records of personnel dosimetry that have been utilized in previous epidemiological studies. Included is information on employment history, radiation exposure, including separate gamma and tritium radiation dosimetry records, as well as vital status information.

We calculate the excess relative rate of leukemia and leukemia excluding chronic

lymphocytic leukemia (CLL) associated with tritium and gamma radiation. The ERR/10mGy (90% HPD) of leukemia associated with tritium and gamma radiation are 0.282 (0.027, 0.678) and 0.044 (0.000, 0.108), respectively. This yields an estimate of the relative biological effectiveness of tritium relative to gamma radiation (RBE) of 6.24 (1.00, 36.09). With regard to leukemia excluding CLL, the ERR/10mGy associated with tritium and gamma radiation are 0.338 (0.048, 0.805) and 0.087 (0.000, 0.195), respectively. This yields an RBE of 3.88 (1.00, 16.80). The values of the RBE are within the range of plausible values suggested by others.

Our results utilize evidence from in vitro and in vivo research to inform estimation of the risk of cancer associated with tritium by incorporating knowledge of the direction and magnitude of tritium's relationship to cancer compared to gamma radiation. This illustrates a simple approach for using Bayesian methods to integrate external knowledge into epidemiological studies without the need to specify estimates of the risk based on research that cannot be easily translated from experimental animal and cellular models into human risk models.

ACKNOWLEDGEMENTS

I am grateful for a number of people without whom this dissertation would not have been possible. First and foremost, my wife, Leah Hope Schinasi, for explaining epidemiologic concepts to me before I even knew I would grow up to be an epidemiologist, and for providing mental support through this process that was much needed. Next, I would like to thank my committee chair, Dr. David Richardson, who has been an indispensible resource of epidemiologic knowledge, academic and mental support, and enabler for my ever growing addition to espresso. In addition, I would like to give special thanks to Dr. Richard Maclehose, who has spent more time than he probably should have explaining Bayesian methods to me. As a small reward, he can rest assured that I have converted to the dark side. I would also like to thank Drs. Steve Wing, Bob Millikan and John Dement for taking part in this lengthy but exciting process.

Additionally, I would like to thank the staff of the epidemiology department, with special thanks to Nancy Colvin for entertaining my never-ending questions. Finally, I would like to thank the baristas of the Global Education Center café, without whom I would have fallen asleep at my desk on countless occasions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CI	Confidence interval
CLL	Chronic lymphocytic leukemia
ERR	Excess relative rate
Gy	Grays
HPD	Highest posterior density
HTO	Tritiated water
LET	Linear energy transfer
LNT	Linear no threshold
LSS	Life Span Study
MCMC	Markov chain monte carlo
RBE	Relative biological effectiveness
SRS	Savannah River Site
Sv	Sieverts
w_f	weighting factor
WBD	whole body dose

CHAPTER 1

INTRODUCTION

1.1 Background

Tritium is a radioisotope of hydrogen that releases β^- energy, which is a form of ionizing radiation. Although there is a volume of research examining associations between ionizing radiation exposure and cancer outcomes, most of this work has been focused on other, higher-energy forms of radiation. For example, an epidemiologically-important study originates from research of the atomic bomb survivors in Hiroshima and Nagasaki, and specifically the Life Span Study of Japanese atomic bomb survivors (LSS). Patterns of exposure in the LSS cohort are distinct from those that are more common in the occupational or environmental setting. In the LSS cohort, individual doses to radiation were mostly in the form of gamma radiation, received in a single event. This is distinct from chronic, low dose exposures that characterize the occupational setting at nuclear fuel facilities.

SRS is unique in that it is the sole nuclear facility in the US with the expressed purpose of producing tritium, which is the primary source of low-energy radiation (specifically, beta radiation) at this facility. As a Department of Energy facility, SRS staff maintained detailed records that have been utilized in previous epidemiological studies. Included is information on employment history, radiation exposure, including separate internal and external dosimetry records, as well as vital status information.

The purpose of the proposed study is threefold. First, we will utilize Bayesian methods to estimate the Relative Biological Effectiveness (RBE) of tritium with reference to gamma radiation. The RBE has been reevaluated by a number of radiation experts, and is the subject of a recent systematic review of the in vivo and in vitro literature [1]. Second, we will utilize the information compiled over the years at SRS to directly address the relationship between leukemia mortality and tritium exposure. Finally, we will present an approach for integrating animal and cellular evidence into epidemiological research in a Bayesian model using informative priors.

This work will inform knowledge of cancer risk due to tritium, a form of ionizing radiation that has been neglected in epidemiologic research, despite its relevance to human populations. In addition, this work may assist the National Institute of Occupational Safety and Health (NIOSH) in decision-making concerning compensation for cancer outcomes among nuclear facility workers.

1.2 Tritium and low energy radiation

1.2.1 Characteristics

Tritium is a radioactive isotope of hydrogen that emits energy, in the form of beta radiation, as it decays into helium [2]. Tritium is created and distributed in a number of ways in nuclear fuel facilities, including neutron irradiation of lithium or deuterium as well as the release of fission-product tritium from reprocessing fuel rods [3-5]. In addition to nuclear facility processes, tritium is created by the interaction of cosmic rays with ^{14}N and ^{16}O, which makes it naturally occurring in the environment [3].

Tritium has a track length of 6 μm, which means it is not capable of penetrating human skin. In its gaseous form, tritium molecules will act as hydrogen molecules and bind to form tritiated water (HTO). Since it acts as water, HTO can enter the body as any other water molecule, through inhalation, ingestion, or absorption through skin [6]. Since tritium can only cause biologically relevant radiation effects upon deposition into the body, the only relevant means of measuring individual dose is via monitoring of body fluids, such as urinalysis [2]. Once absorbed, tritium will diffuse across cellular membranes to uniformly distribute in the body. Tritium has a physical half-life of 12.3 years [7]. ICRP publication 56 includes a two compartment model describing the biological half-life of tritium, which indicates that 97% of tritium is retained in the body for 10 days, while the remaining 3% is retained in the body for up to 40 days [8].

Tritium's short half-life, relative to other radionuclides used in nuclear weapons, means that it requires continued production for use in nuclear warheads. Ongoing

production of tritium means that there continues to be potential for release of tritium into the environment. This makes knowledge about the health effects of tritium pertinent to workers and the public [1]. Relevance to the general public also stems from the fact that tritium is also a waste product of complex power reactors, which continue to be built in the USA.

Tritium is capable of causing cellular level damage by two mechanisms. First, tritium can damage the cell by ionization that releases β^-, or electron, energy. Second, cellular damage can occur when energy is released in nuclear transmutation, or decay, of H^3 to He^3 [2]. This second process releases a positron, or β^+ energy. Malformation and death have been associated with acute exposure to tritium in animal tests [6]. Carsten et al (1989) showed that chronic exposure to tritium at 33 times the maximum permissible concentration (MPC) led to a reduction in marrow stem cells, while a level of exposure 100 times the MPC led to chromosomal damage in liver and micronuclei formation in red blood cells of rats. However, little is known about the effects of chronic low-level exposure. One study by Joksic and Spasojevic-Tisma (1998) found that chromosomal damage to human lymphocytes can result from low-level exposure to tritium[9], but direct estimates of human cancer risk following tritium exposure are not available [10].

1.2.2 Tritium and beta radiation

As mentioned above, tritium emits beta negative (β^-) radiation energy. While some beta emitters will release energy in the form of a high energy positron, tritium emits a negatively charged particle from the nucleus of an atom and is identical to an electron [11, 12]. This negatively charged electron release emits lower energy than a positron and,

in this sense, is more like low energy x-ray and gamma-radiation. Electron tracks from beta radiation are characterized as low linear energy transfer (LET). This is due to their low ionization density compared to forms of radiation like alpha particle radiation [13].

Tritium based beta particles are particularly effective at causing biological damage at the cellular level. First, the average ionization density of high energy x-rays and γ-radiation is lower compared to low energy beta emitters. A higher ionization density means that beta particles can cause more damage within a smaller cellular area than other forms of radiation. For instance, beta particles may be more effective at causing biological changes, such as DNA double strand breaks. This has been shown in Monte Carlo simulations where most double strand breaks that result from low LET radiation are the result of low energy electron tracks occurring together in or near DNA [12]. Low LET electron tracks generate secondary electrons which also have a high ionization density [14].

1.2.3 Relative Biological Effectiveness

Little and Lambert (2008) define the RBE as "the ratio of the absorbed dose of a reference radiation to the absorbed dose of the radiation under consideration that is required under similar conditions to produce an identical level of biological response in a particular animal or cellular study." There are a number of studies examining the RBE of tritium (or beta radiation) in animal and cellular models [15, 16]. However, evaluation of low-level exposure to internal emitters and its relationship to cancer mortality has historically extrapolated from results of the LSS cohort [17]. Researchers have attributed the underestimation of the RBE of beta radiation to the use of high dose rates of reference

radiation. A paper by Ujeno (1983) recommended the use of an RBE of 1 for very large intakes of beta radiation, and a value larger than 1 for environmental exposures to tritium, which would be comparable to chronic, low-level exposure [18]. Another report went as far as to suggest that the RBE of tritium has been underestimated by a factor as great as or over 10 [1].

In a systematic review of the animal and cellular experimental literature concerning the RBE of tritium, Little and Lambert (2008) show that when the results are aggregated, the RBE for tritium with reference to chronic x-ray and chronic γ-ray exposure is 1.19 (95% CI 0.88, 1.49) and 2.49 (95% CI 2.00, 2.98), respectively [1]. The difference between these two estimates is the use of x-rays versus γ-rays as the reference radiation. Since exposures among nuclear facility workers are most commonly forms of γ-radiation, the authors recommend an RBE of 2-3 in studies of health effects of tritium among nuclear facility workers. In a more recent article by Paquet and Metivier (2009), a different approach is suggested for aggregating RBE estimates for tritium. This is based on the type of referent radiation, different exposure scenarios, and the endpoint being considered [19]. The RBE of a nuclide (such as tritium) will appear different according to the level and, perhaps more importantly, the rate at which a dose is administered [20]. In many studies the RBE for tritium is referenced against acute exposures to x-rays and gamma rays. Paquet and Metivier suggest consideration of this information when aggregating an estimate of the tritium RBE [19].

Finally, RBE estimates will depend on the endpoint of interest. The current radiobiological literature concerning tritium's RBE have utilized a range of outcomes, including oocyte survival, leukemia development, thymus weight loss, and chromosomal

damage. Depending on how tritium disperses throughout the body upon ingestion, inhalation, or absorption, the relationship between dose and these separate endpoints will likely vary. For the purposes of application to radiological protection work, concerned largely with cancer outcomes, it will be important to consider only those studies whose endpoints are either some form of cancer or a relevant precursor to cancer development.

1.2.4 RBE versus weighting factor

In epidemiologic studies of the relationship between radiation dose and cancer outcomes, tritium is often aggregated with gamma radiation and examined as 'whole body dose' (WBD) radiation. WBD refers to the fact that tritium and gamma radiation distribute evenly across the body. WBD can be modeled as a single coefficient in excess relative rate or relative risk models. However, the assumption implicit to the current calculation of WBDS is that gamma and beta radiation have an equivalent biological effectiveness. The equivalent biological effectiveness is integrated into radiation protection as a 'weighting factor' that is specified at a value of 1 for tritium. The RBE and weighting factor are related, but used in different contexts.

The 'weighting factor' (w_f) is a mathematical constant that is utilized to quantify the 'effective dose.' The purpose of calculating effective dose is to provide a metric for setting standards for radiological protection. The effective dose represents the sum of the 'absorbed dose' of internally and externally deposited radiation, where the absorbed dose is the energy deposited into body tissue or organs before accounting for different biological effectiveness. WBDS is a measure of effective dose that sums gamma radiation and tritium absorbed doses, where tritium is multiplied by an appropriate w_f.

The effective dose simplifies individual radiation risk assessment by allowing for a single measure of radiation dose [12, 21]. Currently, the ICRP utilizes a w_f of 1 for tritium dose, which states that the sum of the absorbed doses without a w_f is equal to the effective dose, or WBDS, as outlined above [22]. The ICRP asserts that this value is appropriate, based on uncertainties of results from experimental studies of the RBE of tritium and gamma radiation.

The ICRP specifies that the w_f is only intended for use in determining radiological protection standards [23]. Outside of this, ICRP recommends use of the absorbed dose in the organ or tissue of interest when determining quantitative risk in an epidemiological study of the effects of radiation on cancer outcomes. Use of an individual absorbed dose, instead of a summarized 'effective' dose value, allows for estimation of differential effects of multiple forms of radiation with cancer risk. In the case of tritium and gamma radiation, use of individual absorbed dose allows estimation of the RBE for tritium compared to gamma radiation as outlined by Little and Lambert, since the radiation types are not constrained to have an equivalent relationship to cancer risk.

1.3 Radiation Epidemiology and Health Effects of Tritium

There are few studies examining the potential relationship between quantitative estimates of tritium dose and cancer outcomes. Several studies have considered the risk of prostate cancer in relation to tritium exposure status, exposed versus un-exposed, among UK nuclear industry workers [24, 25]. These studies have noted a significantly elevated relative risk of prostate cancer among workers with documented intake of tritium that tended to increase in magnitude with duration of tritium monitoring;

however, tritium doses were not quantified in these analyses and this association could, in part, be confounded by other radionuclide intakes. Thus, it is useful to examine what is known regarding other forms of chronic, low-level radiation dose and cancer among nuclear facility workers. In some studies, tritium doses are either not quantified or were aggregated with external dose, assuming an RBE of 1.

Studies examining the health effects of low dose radiation have been conducted among many occupational cohorts, mostly in the US, Canada, and Europe [26, 27]. In the US, Cragle et al (1988) conducted a study of the SRS cohort in which they found an increased risk of leukemia, excluding chronic lymphocytic leukemia (CLL), but not for 14 other cancer outcomes of interest [5]. Richardson and Wing (2007) conducted an updated analysis of this cohort, in which they considered excess relative risk (ERR) associated with radiation exposure for all-leukemia, leukemia excluding CLL, and myeloid leukemia[28]. Assuming a 3-year lag, and utilizing a nested case-control design, they find an ERR for these outcomes of 4.1 per Sv (90% CI -0.1,11.6), 7.7 per Sv (90% CI 1.4,19.8), and 12.3 per Sv (90% CI 2.1,35.4), respectively.

In the UK, at Chapelcross, McGeoghegan and Binks (2001), found a statistically significant excess relative risk of prostate cancer, but no other cancer outcomes of interest [29]. Though production of tritium is prominent at Chapelcross, production began after 2 of 8 prostate cancer cases left the facility, suggesting tritium is an unlikely cause of the excess risk of prostate cancer. Concerning Sellafield workers, the only statistically significant trend associated with cumulative radiation dose was for leukemia excluding CLL [30]. Tritium specific doses were not taken into account in this analysis.

The study of Canadian workers undertaken by Zablotska et al (2004) found a borderline statistically significant dose-response relationship between radiation and various cancer outcomes, including leukemia excluding CLL (ERR 52.5 per Sv, 95% CI 0.205,291), all leukemia (ERR 18.9 per Sv, 95% CI < -2.08,138) and rectal cancer (ERR 34.1 per Sv, 95% CI 1.41,165), among others [31]. The workers in this cohort have information on tritium dose, but this information is aggregated with external dose information.

Cardis et al. (2007) conducted a pooled analysis of nuclear facility employees from 15 countries [32]. This analysis concluded that workers experience an excess relative risk (ERR) of cancer, excluding leukemia, of 0.97 per Sievert (Sv) increase in radiation exposure. Workers were also found to experience an ERR of 1.93 per Sv for leukemia, excluding chronic lymphocytic leukemia. This study provides useful evidence regarding the relationship between chronic, low dose radiation and cancer mortality; however it does not address the influence that exposure to tritium may have on cancer outcomes.

Little and Wakeford (2008) conducted a systematic review of the epidemiological literature to date concerning tritium. The authors conclude that among those cohorts with tritium-specific dose information, researchers have not examined exposure in a way that would allow for inferences about risk due to tritium [3]. However, tritium dose has been accounted for in studies of nuclear facility workers in the UK, Canada, and US, which the authors identify as the cohorts with the clearest potential for informative research concerning tritium dose and cancer outcomes. Specifically, the authors identify the

Savannah River Site cohort as one of two occupational groups with a large enough pool of workers with "appreciable tritium exposures" that warrant further examination.

1.4 Bayesian Analysis

In attempting to address the relationship of tritium to cancer mortality, one must recognize an inherent lack of power to detect an effect. This is due to two factors. First, individuals typically receive low doses of tritium, resulting in a highly skewed distribution of doses received. Second, our outcomes of interest, for which a relationship to other forms of radiation have been found in the past, are relatively rare. Thus, even a large occupation cohort may have limits in its ability to provide adequate statistical power to detect the relationship of tritium to cancer morality. However, this is further evidence of the need to undertake this research question with Bayesian methods.

Bayesian methods provide the opportunity to combine evidence from radiobiological studies with data from a retrospective occupational cohort study in order to inform estimation of the RBE for tritium and, subsequently, the relationship of tritium and whole body dose radiation to cancer mortality. As discussed earlier, there is a great deal of radiobiological evidence regarding the RBE for tritium, but very little evidence from epidemiological studies integrating or evaluating these estimates. Our study will bridge this gap by drawing on the evidence of the RBE for tritium from results of in vivo and in vitro studies. This information will be used in conjunction with the SRS data and Bayesian statistical methods to provide an epidemiologically based estimate of tritium's RBE.

1.5 Summary

Currently, determination of the maximum amount of tritium exposure a worker is allowed to experience is based on the idea that the biological effectiveness of absorbed tritium dose is equal to that of gamma radiation. There is almost no epidemiological literature directly quantifying the cancer risk associated with tritium exposures. The proposed study will address this knowledge gap using an unusual data resource from an occupational cohort study of workers employed at the SRS, the nation's sole tritium production facility. The fact that most doses received by workers are low makes modeling the relationship between tritium dose and cancer risk difficult, as does examination of a rare outcome. To address this, we utilize Bayesian methods that allow us to integrate knowledge from in vitro and in vivo research into our occupational cohort study of the SRS employees. We illustrate the benefit of using these novel methods for a previously intractable epidemiologic inquiry.

CHAPTER 2

Markov Chain Monte Carlo: A brief introduction for epidemiologists

2.1 Abstract

Bayesian methods offer an appealing alternative to frequentist data analysis that can often handle previously intractable research questions. For instance, Bayesian methods offer a means for epidemiologists to integrate knowledge external to a study. In the case of models with a high number of covariates, highly correlated data, or a multiple comparisons study, a Bayesian approach is often necessary due to computational difficulties with maximum likelihood approaches. The machinery most commonly used for Bayesian analysis is Markov Chain Monte Carlo (MCMC) simulation. Epidemiologists may not utilize MCMC approaches due to lack of familiarity with this procedure and uncertainty about model results. We provide an introduction to the workings of MCMC, and show that, given the same data and model specification, the results of an MCMC simulation match that of standard maximum likelihood estimation (MLE) approaches. In addition, we provide an example of where MCMC provides a clear benefit over MLE data analysis via use of a Jeffreys' prior. We hope this work will make

MCMC less daunting and we look forward to its increased use by epidemiologists in future studies.

2.2 Introduction

Bayesian methods for data analysis are becoming an integral part of the field of epidemiology [33]. Whether the interest is based on its ability to handle complicated epidemiologic problems or the more philosophic foundations of Bayesian methods, such as the belief that integrating prior information is important and useful when conducting data analysis, many epidemiologists see the appeal of these techniques as a complement to frequentist analysis.

Methods for obtaining Bayesian posterior distributions via amendments to frequentist data analysis have been developed and promoted in the epidemiologic literature, the most popular of which is data augmentation. For example, Greenland characterizes data augmentation as the simple inclusion of an extra stratum, or level of information, into an analysis [34, 35]. However, when addressing more complex issues or to arrive at an estimate of the exact posterior distribution, simulation based estimates of Bayesian posterior distributions are often necessary [36].

Markov Chain Monte Carlo (MCMC) methods are the most common machinery for simulation based estimation of Bayesian posterior distributions. Despite its accessibility in many software packages, including WinBUGS and Statistical Analysis Software (SAS), few epidemiologists currently utilize MCMC procedures. This is likely due to two factors. First, simulation based methods for obtaining the distribution for parameters of interest are unfamiliar. Transitioning from maximum likelihood to

simulation based methods requires some explanation of their similarities and differences. Second, identifying convergence of MCMC simulations is quite different from identifying convergence of an algorithm to find a maximum likelihood estimate. We must be comfortable reading visual output that indicates whether or not model results can be considered reliable.

In this article, we will address both of these issues. First we will discuss MCMC and offer a brief overview of Monte Carlo integration and Markov Chains. Next, we will provide guidance for epidemiologists on the use of MCMC procedures and illustrate how to evaluate model convergence. We will illustrate how MCMC and frequentist regression models are similar, and when MCMC provides a clear advantage over standard maximum likelihood approaches for estimating regression model parameters. We focus on 3 visual plots that are standards for assessing MCMC model convergence and offer some guidance on how to read these plots and use them to avoid some potential pitfalls in MCMC analysis. For simplicity, we will focus on analysis of a binary outcome variable using a standard unconditional logistic regression model with a single explanatory variable. We will assume a basic understanding of Bayesian probability theory, which has been discussed elsewhere [37-40]. We hope this discussion will ease the path to MCMC for epidemiologists.

2.3 Monte Carlo Integration

Epidemiologists routinely use maximum likelihood regression methods to derive parameter estimates. Given the observed data, a regression model implies a particular likelihood function; an estimate is chosen for the parameter of interest such that it

maximizes the observed likelihood, hence the name maximum likelihood estimate (MLE). Despite simple conceptualization, obtaining an MLE can be a complex process that depends on the ability of an algorithm to converge at some estimate of the parameter of interest, given the specified model and data. In many situations, it is relatively easy to obtain model convergence, although for some regression models, such as log-binomial regression convergence, it may be difficult. In addition, MLE may fail when models have a large number of covariates, highly correlated covariates, or even few covariates but multiple effect measure modifiers. In these cases, where maximization of a likelihood function is difficult, epidemiologists may benefit from alternative methods for obtaining parameter estimates for their regression models.

Monte Carlo integration allows researchers to handle epidemiologic inquiry where maximization of the likelihood is unachievable because it does not require the maximization that can pose a challenge in MLE. Suppose that, regardless of whether our inquiry is Bayesian or frequentist in nature, that we are interested in a model for the expectation of some function of the random variable X, denoted $E[f(X)]$. The random variable could be the probability of the outcome and the function a log odds or some other effect estimate of interest. *If* there exists a distribution for the random variable of interest, and *if* we can draw 'n' random samples from the distribution of X, that as $n \to \infty$, the average of f(x) over those samples will converge to the value of interest. This can be written as follows:

$$E[f(X)] \approx \frac{1}{n}\sum_{i=1}^{n} f(X_i) \quad . \qquad (1)$$

Thus, we can estimate the expected value of some function of a random variable by repeated sampling.

Use of MC integration does not necessitate a major departure from evaluation of parameter estimates with which epidemiologists are most familiar, such as the 95% confidence interval (referred to in Bayesian analysis as a 95% credible interval). As implied by equation (1) we can use simple summary statistics (such as the median and percentiles of the distribution) to provide a parallel to the point estimates and CIs which we would obtain with standard regression modeling. If we know that a parameter is normally distributed with mean μ and variance σ^2, we can easily compute a 95% CI around the effect estimate of interest. The same calculation can be done with Monte Carlo integration. Say we draw 1,000 random samples from the normal distribution. The 2.5^{th} and 97.5^{th} percentile values will correspond to the limits of a 95% confidence interval, and the median will equal the point estimate of interest.

The definition of Monte Carlo integration contains a big 'if.' That is to say, 'if' we can identify a distribution from which we can draw random samples, we can be confident that our Monte Carlo integration will accurately summarize the characteristics of that distribution. In practice, it is Markov Chains that allows us to identify this distribution.

2.4 Markov Chains

Markov chains provide a way to sample from the posterior distribution that is the focus of a Bayesian inquiry. Once the chain finds the posterior distribution (known as the

stationary distribution of the chain) we can draw as many samples from it as we like and summarize them via Monte Carlo integration to obtain parameter estimates of interest.

A Markov chain can be thought of as directed random walk through the parameter space (that is, all the possible values of the parameter of interest). Some values in the parameter space are more likely than others (given the prior distribution and the observed data) and a well-constructed Markov Chain sampler will sample from these 'more likely' regions of the sample space. In fact, a well-constructed Markov Chain sampler will eventually draw random samples from the posterior distribution of interest, regardless of where the chain is started (although in practice, the eventuality of this might take quite a long time to be realized). Many statisticians describe this as the Markov Chain 'forgetting' the initial values it draws. In other words, a Markov Chain should identify a posterior distribution of interest regardless of the starting point of the chain. This notion is used in the Gelman and Rubin diagnostic which relies on running parallel chains that begin at different initial values and checking to see that they converge at the same posterior distribution. We utilize this diagnostic in our example 2.

The technical details of Markov Chains can be daunting and there are many algorithms that produce these 'random walks' through the parameter space. The default method for WinBUGS and SAS is the Gibbs sampler. However, there are also other algorithms, such as Metropolis-Hastings (of which Gibbs sampling is a special case), slice sampling, or Langevin-Hastings. For those wishing for further detail, we recommend the seminal paper by Gelfand and Smith, as well as the texts by Carlin and Louis, Gilks and Richardson, or Gill [39, 41, 42].

2.5 Markov Chain Monte Carlo estimation

MCMC offers an approach to estimation of parameters in a regression model, much like maximum likelihood methods. MCMC is a potentially useful complement to the standard tools used by epidemiologists for estimation of associations, albeit one that is often more computationally-intensive than maximum likelihood approaches.

An important point for epidemiologists to understand is that, in the presence of a diffuse (non-informative) prior distribution and a well identified linear model, the posterior distribution for a model parameter obtained using MCMC methods will approximate the MLE, which says the data and the model specification drive the entirety of the estimation process. In other words, the parameter estimates of a Bayesian and frequentist model will be, essentially, the same. In terms of Bayes' theorem, one might consider the use of a non-informative prior to look like this:

$$P(\theta|D) \propto P(\theta)P(D|\theta) = P(D|\theta)$$

where $P(D|\theta)$ is the likelihood. Here, since $P(\theta)$ does not contribute any information to the model, the posterior distribution is proportional to the likelihood of a frequentist analysis.

One situation in which MCMC is a potentially useful complement to standard maximum likelihood approaches used by epidemiologists for regression model estimation is when there are sparse data. For example, standard logistic regression provides estimates of the probability of an outcome that are bounded at 0 and 1. This is an appealing aspect of logistic regression since the actual probability of any outcome can

never be less than 0 or greater than 1. This is most helpful when the probability of the outcome for an individual covariate pattern are somewhere in the middle of this range, so values greater than 0 and less than 1. However, when the probability of the outcome in a specific stratum of a predictor variable is exactly 0 or 1, a logistic regression model will produce estimates that are biased away from these extremes. A simple conceptualization of this is to think of a 2x2 table where one of the cells contains no observations. In epidemiology, this is referred to as nonidentifiability. Another term used by some is separation [43].

This event can be thought of as an extreme case of small sample bias. In logistic regression, it is unique because even though the parameter estimates and associated errors may be unreliable, a standard logistic regression model will still obtain model convergence, with no warnings about reliability of estimation. This convergence may suggest that the parameter estimates are reasonable when, in fact, they are not. MCMC offers an approach to diagnosing and resolving such convergence problems.

2.6 Diagnosing Model Convergence in MCMC

Assessing model convergence for MLE using standard regression techniques is a simple yes/no diagnostic. In fact, convergence of a model happens in the background using a pre-specified optimization algorithm (such as Newton-Raphson) and typically requires little attention from the researcher. However, convergence of an MCMC procedure requires the user to evaluate several diagnostics, the most common of which are visual summaries. Three visual plots are routinely used to assess model convergence:

1) Trace plots: these plots document the sample drawn (y-axis) at each iteration (x-axis) of the MCMC procedure. Once the chain has identified the stationary distribution, these samples will appear to be randomly sampled from a narrow region of the y-axis that will correspond to the general region of the posterior density of interest.
2) Autocorrelation plot: these plots document the correlation (y-axis) of samples at each step of the chain with previous estimates of that same variable, lagged by some number of iterations (x-axis). Ideally, the autocorrelation declines rapidly so that, eventually, we can be confident that the samples from the stationary distribution can be thought of as random, and not reliant on initial values from the chain.
3) Density plot: as its name implies, this plot is a summary of the sampled values that define the stationary distribution, which approximates the posterior distribution of interest. The peak of this density will be the point estimate, or the value with the most support from the data and specified prior. Kernel density estimation is used to smooth over the samples and produce an estimate of the posterior distribution.

Utility of the visual plots depends on the MCMC process having been run for a sufficient number of iterations. There is no rule for the number of iterations necessary to achieve model convergence. This value varies across datasets and model specifications. However, as a general rule, running a chain for longer will not negatively impact the results.

In addition to the visual plots discussed above, a Gelman Rubin diagnostic check is often used to assess model convergence. This diagnostic involves starting the MCMC process at three different starting values. If the models do not obtain the same estimate for a parameter of interest, this is evidence that the process is sensitive to the starting point, and that the model has not converged. It should be noted that small differences in the estimates may occur if each model is not run long enough to minimize simulation error.

2.7 MCMC Approaches to Addressing Problems of Nonidentifiability

Many solutions to problems of nonidentifiability have been proposed. Subjective priors are a particularly appealing approach when subject matter exists to inform that prior [44, 45]. Other times, researches may feel somewhat unsure about incorporating substantive information into an analysis. In this case, so-called objective Bayesian approaches may be appealing. Various objective priors have been proposed in the statistics literature [46, 47]. A common objective prior is Jeffreys'. Although Jeffreys' prior is typically referred to as a non-informative prior, this name can be somewhat misleading from a substantive point of view. In some models (for example, a linear regression) Jeffreys' prior is a uniform distribution and is truly uninformative, in the sense that it conveys no information about the prior belief in the magnitude of the effect of interest. In other regression settings this is not the case. As shown by Ibrahim et al and Chen et al [48, 49], the Jeffreys' prior in a logistic regression is a weak symmetric prior centered at 0 that has somewhat heavier tails than a normal distribution and lighter tails than a t-distribution. The Jeffreys' prior in a logistic regression conveys some prior

information, but not much. However, this tiny amount of prior information can be enough to allow for easier estimation of effects in situations with non-identifiable parameters.

2.8 Example

We illustrate the use of these diagnostic plots in the context of a simple logistic regression model. The data we use came from the 2008 NC live birth records collected by the North Carolina State Center for Health Services (SCHS). We constructed a case-control dataset where cases are defined as newborns with low birth weight (<2500 grams) among full term births ($\geq$ 37 weeks of gestation), and controls are all other full term births. In addition, we restrict the population to mothers younger than 28 years. This yields 4,412 cases and 5,019 controls. We examine the relationship between maternal education dichotomized as those having less than college education (maternal education <13 years) compared to those having any college education (maternal education $\geq$13 years) and term low birth weight. Our model is specified as: $\text{Logit}(P(D=1)) = \beta_0 + \beta_1 x$, where 'x' represents our explanatory variable of interest, maternal education (0 if less than college education, 1 if any college education). We also fitted a regression model with 3 categories of maternal education: no education, 1 year of education, and >1 year of education (the reference category). This example is presented because it provides a simple illustration of a case of separation in logistic regression. In addition, since the idea of an outcome of low birth weight among full term births among mothers with 1 year of education is not a biologically implausible outcome, the idea of addressing non-identifiability with a statistical correction may be appropriate. We conduct these analyses

in PROC MCMC and PROC GENMOD using the BAYES statement in SAS (V 9.2 Cary, NC) to provide the examples that follow. These models were run for 100,000 iterations with a 10,000 iteration burn-in period. Initially, the specified priors for all regression model parameters were diffuse, i.e., N ~ (μ=0, σ^2=100,000. In addition, a model was fitted specifying Jeffery's prior using SAS PROC GENMOD (coeffprior=jeffreys in the BAYES statement). To provide a parallel to the parameter estimate and standard error obtained in a frequentist model, we will report mean values and standard deviations for the Bayesian results.

2.9 Results

We first fit a model with a single binary explanatory variable. The frequentist and Bayesian models return similar results. For β_0, the frequentist and Bayesian model estimates are -0.156 (std error: 0.025) and -0.156 (std error: 0.025), respectively. For β_1, the frequentist and Bayesian model estimates are 0.082 (std error: 0.047) and 0.082 (std error: 0.047), respectively. The diagnostic plots produced by this model provide an example of good MCMC model convergence. Figure 2.1 documents these plots. An MCMC algorithm that has converged will tend to wander randomly around the same area, rarely venturing outside that area. The trace plot is difficult to read, but it suggests that the chain is wandering through the same region of the parameter space. This plot suggests that the chain has found the stationary distribution, but it is not guaranteed. For this reason, we must use this in conjunction with the autocorrelation and density plots to confirm model convergence. The autocorrelation drops precipitously, from lag 0 to lag

50, which tells us that each sample in the chain is only slightly correlated to the previous draw.

Finally, the density plot allows us to assess what the distribution around that point estimate is, obtained via kernel density estimation. In this case, we see a symmetric distribution around the parameter estimate of interest. The density plot is most useful when assessing whether or not a chain started from different points converges to the same distribution. This is essentially a visual Gelman-Rubin diagnostic.

When we treat each level of maternal education as a single category, the stratum of mothers with 1 year of education contains no cases of low birth weight among term births, but contains 9 controls. From a standard logistic regression, the ln(odds) comparing mothers with 1 year of education to mothers with any education greater than 1 year is -20.237 (standard error: 11824.66). The odds ratio based on this value is 2.1×10^{-9}. Clearly, logistic regression is not providing an informative effect estimate for this relationship, but since its probability is within the range of possible, though not plausible, values (i.e., greater than 0), model convergence is achieved.

When using MCMC methods to derive estimates for the model parameters, we do not report an estimate of the ln(odds) for this parameter here because, when started from three different points, the MCMC samples do not seem to be coming from the same part of the parameter space. This strongly suggests non-convergence, but it is instructive to consider the diagnostic plots from one set of these results. Figure 2.2a documents the diagnostic plots of this estimation and provides an example of non-convergence of an MCMC procedure. The trace and autocorrelation plots are particularly informative here.

Compared to the trace plot in figure 2.1, which shows a dense area from which the MCMC procedure repeatedly selects independent and identically distributed, iid, samples, we can see the Markov chain trying to find the stationary distribution, but never settling into an area from which it can draw independent samples. It appears as though estimates are serially correlated to previous estimates in the chain. This is confirmed by the autocorrelation plot, which never approaches 0. The kernel density plot confirms what we already know, that the chain was unable to find a smooth distribution with a single point estimate with more support than any other. For simplicity, we only present plots from the first of three MCMC procedures run from different starting points. However, we should point out that the results for the other procedures show a similar lack of model convergence.

The final model using Jeffrey's prior was run for 100,000 iterations with a 1,000 iteration burn-in. When we specify the Jeffrey's invariant prior into the same model described above, the ln(odds) calculated is -3.3254 (std dev: 2.0076). Figure 2.2b documents the diagnostic plots for estimation of this parameter. Here, we see a series of diagnostic plots that look more like figure 2.1 that figure 2.2a. In other words, use of the Jeffrey's prior results in MCMC model convergence.

2.10 Conclusion

In this article, we have provided epidemiologists with a simple introduction to Markov Chain Monte Carlo methods for statistical inference. First, we showed that, given the same information, a Bayesian and frequentist model will calculate the same parameter estimates. We did this analysis to make clear the point that inference via MCMC

simulation is not necessarily different from standard regression techniques currently used by epidemiologists. Once this point was made, we showed an example of a case where Bayesian modeling provides a clear advantage over frequentist modeling via simple implementation of a correction for separation in logistic regression, which is unavailable in frequentist modeling procedures. However, the advantages of Bayesian modeling are not limited to handling inappropriate model results. In its simplest form, Bayesian modeling is a learning process that allows the researcher to combine prior evidence of an exposure-disease relationship with that evidence that exists in the data at hand. Clinicians, in particular, may find this appealing in terms of decisions related to whether or not a certain drug is appropriate for a given patient [50].

One limitation of Bayesian modeling is the computational power required for some models. In the case where a researcher is interested in a model with many covariates, or a large dataset, an MCMC procedure may take hours, even days, to complete. This is something that may be overcome with improvements in computer technology, but may be off-putting to some. Additionally, it is worth noting that in many cases where there is no prior information of interest to a researcher, a frequentist analysis will very likely yield the same results as a Bayesian analysis in much less time.

As with any valid epidemiologic inquiry, the model specification must be correct. The formulation of epidemiologic inquiry and correct model development are discussed elsewhere [51-53]. In the context of MCMC, if a model is not properly specified, the results will be incorrect, even if a Markov chain converges. The Markov chain may not converge in the presence of improper model specification, but this is not guaranteed.

Additionally, the more complex the model, the more difficult it becomes to diagnose proper model convergence. Thus, care must be taken in developing the correct model.

An additional concern, mentioned earlier, is that there is no established method for deciding what an appropriate number of iterations and burn-in might be. Rather, we use a trial and error process, where the ultimate goal is to obtain stable parameter estimates that minimize simulation error. Like the computational intensity discussed above, this will require more time on the part of the researcher. However, Bayesian methods are indispensible as a tool for handling intractable epidemiological inquiries.

We want to emphasize that these methods are not limited to use in obtaining Bayesian posterior distributions. Proponents of frequentist statistics can just as easily use the machinery of MCMC to obtain parameter estimates of interest. We hope that readers will find MCMC less daunting, and we look forward to its increased use by epidemiologists in future studies

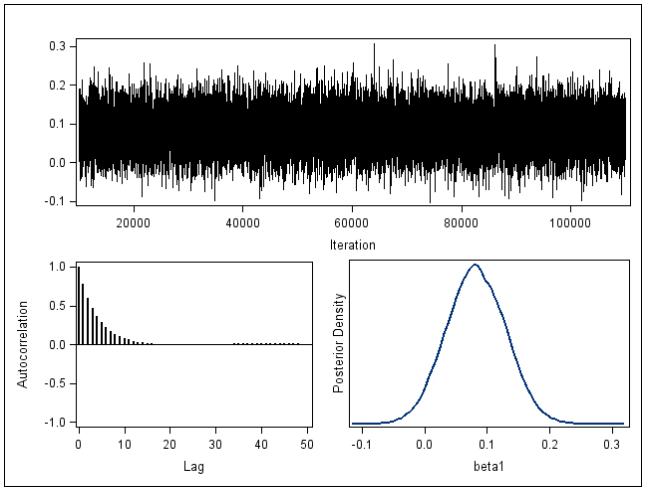


Figure 2.1. Diagnostic plots for estimation of β_1 in example 1.

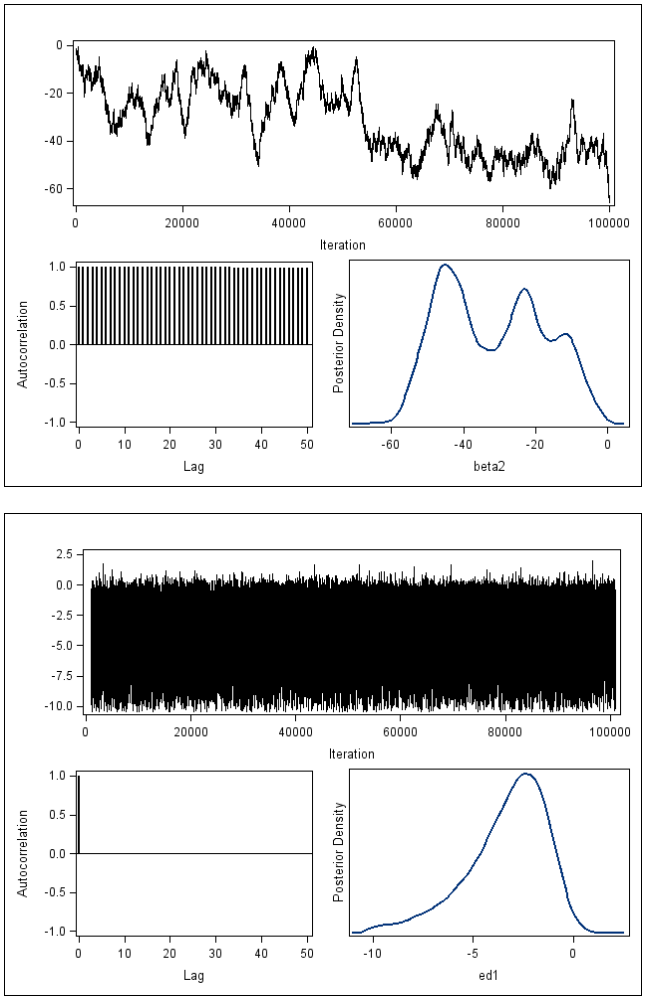


Figure 2.2. a) Diagnostic plots showing non-convergence of an MCMC procedure (top). b) Diagnostic plots showing convergence of the same model using a Jeffreys' prior.

Chapter 3

Informative priors: a simple approach for utilizing animal and cellular evidence in observational research via order constrained priors

3.1 Abstract

Informative priors can be a simple and useful tool for epidemiologists to handle sparse data problems in regression modeling. It is sometimes the case that previous research may indicate the direction of effect or magnitude of one exposure's effect relative to another but may not provide enough information to specify the distribution of a prior in terms of the magnitude of a parameter describing an exposure-disease relationship in an epidemiological study. When considering human carcinogens, for example, an important source of knowledge is derived from toxicological studies and experimental research. Incorporating this knowledge as a prior in regression analysis of epidemiologic data often is difficult since the findings cannot be considered exchangeable across species or from cellular level outcomes to mortality and morbidity. We present a method to help bridge the gap between animal and cellular studies and

epidemiological research by truncation and specification of an order constrained prior, illustrating how external information from toxicological and experimental research regarding parameter associations may be usefully incorporated.

3.2 Introduction

Epidemiologists face a number of challenges in developing etiologic models of disease. When attempting to develop a model for a specific exposure-disease relationship, adjusting for confounders and effect measure modifiers of interest, we often deal with data that suffer from a lack of information and do not allow us to develop the models we believe to be most appropriate. This lack of information is often in the form of sparse data, which is a form of information bias [54]. In its most extreme form, where we have no information for a specific covariate pattern, this bias will be a case of non-identifiability [55]. Frequentists will approach this problem through model building exercises, where potential confounders and effect measure modifiers are systematically introduced or removed based on various rules aimed at balancing the tradeoff of bias introduced with precision gained [52]. Bayesian methods provide an approach to investigating associations in settings of sparse data, which may arise when models include many covariates, effect measure modifiers, or with low or rare exposures [41, 42]. One such technique is the use of informative priors. Informative priors can be thought of as external knowledge integrated into an analysis that will influence parameter estimation when the epidemiologic data are sparse. Although the benefits of informative priors have been well documented in the statistical and epidemiologic literature they are rarely utilized [44, 56]. However, they can be a particularly useful tool for epidemiologists who are facing challenges in model development and data analysis.

In this paper, we will discuss the history and challenges of informative priors in epidemiological literature. Next, we will introduce a simple approach to incorporating an informative prior that may be useful for epidemiologists. This involves a priori specification of a model parameter based on evidence from experimental animal and cellular research into epidemiologic studies using an order constrained prior via truncation. Order constrained priors are used in dose-response analyses to improve estimation of parameters for categories of exposure based on the proposed direction of the relationship between that exposure and disease. For instance, if 3 levels of exposure are specified from lowest to highest as, β_1, β_2 and β_3, we constrain a parameter estimation so that $\beta_1 \leq \beta_2 \leq \beta_3$. We will illustrate two examples of this approach. The first uses evidence regarding the relationship between radiation exposure and cancer risk. The second is from an analysis of the relationship between exposure to tritium and leukemia outcomes. These examples will show how integrating truncation or an order constrained prior can provide a benefit to an epidemiologic analysis without the need for assigning an absolute value to the prior distribution of exposure-disease relationship, as is often done in Bayesian analysis. Although the researcher utilizes the machinery of Bayesian analysts, we believe this is a simple extension of an established approach for informing parameter estimation, which some researchers may find appealing.

3.3 Methods

The benefit of informative priors is most clear in cases where sparse data leads to problems in regression modeling. A good example of this is the case of nonidentifiability, which can occur when a specific covariate pattern contains no observations [55]. One

way to conceptualize this is to think of a 2x2 table with either no exposed cases or no unexposed controls. In logistic regression models, this will lead to separation, where estimates or a parameter of interest will approach +/- infinity [57]. In this situation, a Jeffrey's prior can be utilized to stabilize parameter estimates, as demonstrated by Heinze et al [43]. The appeal of a Jeffrey's prior is that is it a data-driven prior, and does not require the researcher to specify an a priori distribution for any parameters of interest.

Another informative prior which is based on ideas about the change in magnitude of an effect is implemented through specification of a Cauchy distribution, recommended by Gelman et al. [44]. This prior is specific to logistic regression, and simply presumes that any given input variable is unlikely to lead to a change of 5 on the logistic scale (or, for example, a change of the predicted probability from 0.01 to 0.50) [44]. The authors recommend use of a Cauchy distribution as a 'default' prior to overcome small sample bias in logistic regression. The use of the Cauchy distribution is built on a history of epidemiologic evidence where, typically, major changes in the predicted probability of an outcome of interest associated with any individual covariate are very rare. This approach is similar to that suggested by Greenland where priors are specified based on 'weakly informative ranges' for the expected relative risk [58].

As shown above, the more popular way to integrate an informative prior into an analysis is to identify a means of allowing the amount of data available to drive the importance of the prior. This is supported by Gelman et al.'s recommendation that the Cauchy distribution be used as a 'default' prior in any logistic regression model. However, it is possible to integrate informative priors for a parameter of interest based on existing knowledge without explicating the distribution of that parameter. We suggest the

use of experimental animal and cellular evidence to direct estimation of coefficients of interest when investigating questions regarding environmental or occupational hazards for which one can specify an order to direction of effect.

3.3.1 Prior knowledge based on experimental data

Experimental animal and cellular studies can provide useful information to epidemiologists regarding specific exposure-disease relationships of interest. They benefit from a controlled environment, which minimize the biases that typify epidemiological research. With regard to exposure or dose, the researcher has some control and knowledge of the amounts received by an animal or cell that is often more accurate than estimates evaluated in epidemiological research [59]. Additionally, these studies can be conducted in order to inform etiologic relationships that are unethical for controlled studies in humans, such as exposure to radiation and cancer outcomes.

Applying the evidence from experimental animal and cellular models to inform epidemiological epidemiologic research is difficult. For example, an exposure's relationship to disease may be different for humans compared to other animals. In other words, parameter estimates are not necessarily exchangeable across species. A similar limitation for cellular models is that they are often concerned with disease precursors, which limit their ability to inform disease outcomes. This is common in cancer research, where an experiment will evaluate endpoints such as chromosomal aberrations instead of cancer incidence or mortality.

Even if a researcher can reliably translate the direction and magnitude of an exposure's relationship to disease from an experimental animal to human model, the

statistical uncertainty around the parameter estimate of interest will often be overly precise. This is due to the fact that experimental animal and cellular studies easily benefit from a large number of observations. The researcher often has the ability to simply increase the number of observations to obtain statistically significant results. In order to apply evidence from these studies as prior information in an epidemiological study, a researcher may simply multiply the standard error associated with a parameter of interest by some factor that is meant to represent the increase in uncertainty associated with moving from an experimental animal or cellular to human cancer risk model [60, 61]. However, this conversion is arbitrary and not necessarily based on any established metric for applying evidence across species.

We propose an alternative approach for using evidence gained from these studies as prior knowledge in a Bayesian analysis of epidemiological data in a way that does not require assuming a high degree of exchangeability for a parameter of interest, given the limitations mentioned above. We suggest the use of an order constrained prior via truncation. Order constrained parameters have a history of use in the dose-response literature where, as the level of exposure increases the effect of that exposure on the outcome of interest is specified to be greater than the level of exposure preceding it [62].

When attempting to inform an epidemiological analysis with results from experimental animal and cellular research, an order constrained prior provides an intuitive way of integrating results that avoids the pitfalls of trying to directly apply effect estimates across species or outcomes. We present two examples of this approach. First, we show an example where experimental evidence of the direction of the relationship between cancer risk and radiation exposure can inform estimation of the leukemia risk

due to whole body ionizing radiation (WBD) in an epidemiological study via truncation. Second, we use of an order constrained prior to estimate the excess relative rate (ERR) of leukemia by integrating evidence of the relative biological effectiveness (RBE) of tritium referenced against gamma radiation.

3.3.2 Linear no-threshold models

In order to provide a context for the examples below, we briefly discuss linear no-threshold models and RBEs in radiation epidemiology. First, in studies of nuclear fuel facility workers exposed to ionizing radiation, estimation of the relationship between radiation and cancer outcomes utilizes a linear-no threshold (LNT) model. This model has two characteristics: first, it suggests that that the relationship between chronic, low dose radiation exposures and cancer is monotonic. Second, the LNT model assumes that there is no level of radiation that can be considered as 'safe' or inherently unable to cause cancer outcomes, as is the case with the threshold dose-response model. This latter characteristic of the LNT model is supported by the experimental animal and cellular literature [11]. In brief, the stochastic nature of cancer induction tells us that a single mutation or DNA strand break is capable of causing the events that lead to cancer. It follows from this that exposure to radiation does not afford an individual any immunity to cellular level damage. Since even very low doses of ionizing radiation may initial the events that lead to cancer, the relationship between ionizing radiation is either a positive relationship or a null relationship. In other words, the slope of a parameter estimating a linear relationship between radiation exposure and cancer is greater than or equal to 0.

3.3.3 Relative biological effectiveness

Little and Lambert define the RBE as "the ratio of the absorbed dose of a reference radiation to the absorbed dose of the radiation under consideration that is required under similar conditions to produce an identical level of biological response in a particular animal or cellular study" [1]. In a regression model, this can be thought of as the ratio of the beta coefficients (or slope) for two forms of radiation in a model of cancer outcomes. Little and Lambert conducted a systematic review of animal and cellular studies and showed that for a number of different cancer precursors and endpoints, the biological effectiveness of tritium, an emitter of β^- radiation, is greater than that of gamma radiation. This conclusion is based on the fact that although the magnitude varies across endpoints and study designs, the direction of the relationship between tritium dose and the outcome of interest is also of greater magnitude than gamma radiation dose.

3.4 Example

The experimental research provides us with a strong expectation for the relationship between cancer risk and radiation exposure, but no intuitive way of using this knowledge. With regard to the LNT model, we know that the relationship between radiation dose and cancer is either null or positive. This can be built into modeling via truncation of a slope coefficient at a lower value of zero. We illustrate this by analyzing the relationship of effective whole body dose and leukemia, comparing these results to a model without this truncation. With regard to the RBE evidence, we expect that the risk of cancer associated with tritium will be greater than or equal to that of gamma radiation. We illustrate this by modeling the relationship of absorbed tritium dose and gamma

radiation dose with leukemia. We utilize data from an occupational cohort of nuclear fuel facility workers at the Savannah River Site in Aiken, SC.

Dose-response relationships are estimated using an excess relative rate (ERR) model in two forms:

Model 1: $RR = e^{\alpha i}(1+\beta*(d_g + d_t))$

Model 2: $RR = e^{\alpha i}(1+(\beta_1*d_g + \beta_2*d_t))$

where '$e^{\alpha i}$' indexes the baseline rate within strata 'i' for the matching variables attained age, sex, race, pay code, birth cohort, and employment status at the end of follow-up. β represents the ERR for a given exposure, 'd_g' represents cumulative dose from gamma radiation, and 'd_t' represents cumulative radiation dose from tritium. In model 1, the ERR is estimated for the unweighted sum of tritium and gamma radiation dose, which represents cumulative whole body dose, measured in mSv. We consider a truncated prior for this model where $\beta \geq 0$. In model 2, separate terms are considered for cumulative gamma dose and tritium dose in mGy, a metric for absorbed radiation dose. In model 2, we consider an order constrained prior in the form $\beta_2 \geq \beta_1$. All doses are lagged 3 years from the time of leukemia death for cases and time of selection into a risk set for controls. The distributions of cases and controls across matching factors are presented in Richardson and Wing [28].

We implement this model using the Markov Chain Monte Carlo procedure (PROC MCMC) in the SAS software program (v9.2 Cary, NC). In all models, we assign prior values for the parameter estimates of interest such that the shape of the posterior

distribution is not informed, or $\beta \sim N(\mu=0, \sigma^2=100{,}000)$. In our truncation prior example, estimation of WBD is specified so that $\beta \geq 0$. Our order constrained prior for β_2 includes specification that the slope of the coefficient for the tritium dose effect is greater than or equal to the slope of the coefficient for the gamma dose effect, or $\beta_2 \geq \beta_1$. In this case, the specification of the prior distribution for β_1 remains $\beta_1 \sim N(\mu=0, \sigma^2=100{,}000)$, while we specify $\beta_2 \sim N(\mu=\beta_1, \sigma^2=100{,}000)$. Because MCMC utilizes repeated sampling to identify the posterior distribution for a parameter of interest, including this truncation for β_2 as stated above restricts parameter estimates at each iteration of the Markov chain so that the sampled value of β_2 cannot be less than the sampled value of β_1.

All models were run 3 times for 100k iterations with a 10k burn-in and 3 distinct starting values to emulate a Gelman-Rubin diagnostic check to ensure that results are not sensitive to starting values of β_1 and β_2. Final models were run for 10 million iterations with a burn-in of 100,000 iterations to minimize simulation error. Effect estimates are reported as the median of the distribution for parameters of interest. For consistency with the existing radiation epidemiology literature, we report 90% highest posterior density (HPD) intervals. Since we are using a linear relative rate model where bounds of the HPD can cross 0 to include negative values, we calculate confidence limit ratios (CLRs) as the upper minus the lower bound divided by 2.

Table 3.1 presents results from truncation of WBD. When estimating the relationship between Leukemia excluding CLL, the precision and estimate of the ERR/10mSv of whole body dose are almost identical. This is due to the fact that 90% HPD in an unconstrained model does not include any values below 0. Regarding the

models estimating the ERR/10mSv of whole body dose, the 90% HPD of the non-truncated model includes negative values and has a CLR of 0.063. In the constrained model, these values are eliminated, leading to a modest improvement in the precision, shown by a reduction in the CLR to 0.059.

Table 3.2 presents MCMC analyses of leukemia and leukemia excluding CLL. When the model includes a non-informative prior for both exposure and no truncation, the estimate of the ERR/10mGy due to gamma radiation (β_1) for leukemia and leukemia excluding CLL are 0.053 (90% HPD: -0.025, 0.142) and 0.176 (90% HPD: 0.011, 0.375), respectively. The estimated ERR/10mGy due to tritium (β_2) for leukemia and leukemia excluding CLL are 0.141 (90% HPD: -0.323, 0.649) and -0.281 (90% HPD: -1.136, 0.548), respectively.

When the model simulation is restricted so that $\beta_2 \geq \beta_1$, retaining a non-informative prior for β_1, estimates of the ERR/10mGy due gamma radiation for leukemia and leukemia excluding CLL are 0.034 (90% HPD: -0.031, 0.110) and 0.082 (90% HPD: -0.017, 0.206), respectively. The estimated ERR/10mGy due to tritium for leukemia and leukemia excluding CLL are 0.298 (90% HPD: 0.027, 0.702) and 0.344 (90% HPD: 0.049, 0.817), respectively.

To provide a metric for precision of the estimates, table 3.2 includes confidence limit ratios (CLR) based on the width of the confidence bounds divided by 2. The CLR for all effect estimates and outcomes decrease when comparing the model including truncation to the model with a non-informative prior specified for β_1 and β_2.

3.5 Discussion

We have presented a method for integrating an informative prior that bridges the gap between experimental animal and cellular studies and epidemiological research. The appealing aspect for epidemiologists attempting to apply evidence from animal and cellular models is the fact that this approach avoids the need for specifying a prior distribution for the parameter of interest.

In the example of truncation of an LNT model at a lower bound of 0 based on evidence of radiation as a carcinogen, we see a modest change the model results. This is due to the fact the bulk of the posterior density was already above a minimum bound of 0.When moving from a model with no prior to a model with truncation of $\beta_2 \geq \beta_1$, the precision of all model parameters improve. The parameter with the clearest gain from the use of truncation is the estimate of the relationship between tritium and leukemia excluding CLL. When we truncate estimation of tritium in this model the estimate shifts from an implausible negative value to a positive value. Table 1 also shows that the ratio of the confidence bounds is less than half of what it is in a model excluding truncation.

In addition, the results including truncation for leukemia excluding CLL are similar for leukemia as the outcome, indicating an improvement in model fit. This is because we expect the parameter estimates for these two outcome groups to be somewhat similar, but slightly biased for leukemia including CLL due to a longer time course for CLL versus other forms of leukemia. Also, since it is known that radiation exposure is not protective for leukemia, or any form of cancer, we believe this shows a clear example

where sparse data has led to a biased result and, thus, a case where an informative prior can be very useful.

Similar to the vaguely informative priors recommended by Greenland and Gelman [44, 54], we believe the order constrained prior to be a useful tool that can be implemented when the researcher has some knowledge concerning the direction and magnitude regarding a parameter estimate of interest relative to another parameter in the same regression model. An appealing aspect of truncation as an informative prior is that the researcher is not required to take part in synthesize evidence from multiple sources to calculate a prior distribution for any parameter in the regression model. Rather, simple knowledge of the direction of effect of one parameter compared to another in the same regression model is sufficient. Thus, implementation of this method is straight-forward, and we hope that this approach will be appealing to Bayesians and frequentists alike.

Table 3.1 Parameter estimates for the ERR/10mSv of whole body dose radiation.

Outcome	Truncation	Parameter Estimate (90% HPD*)	CLR
Leukemia			
	$\beta \geq 0$	0.055 (0.000,0.117)	0.059
	none	0.054 (-0.003,0.123)	0.063
Non-CLL Leukemia			
	$\beta \geq 0$	0.099 (0.010,0.207)	0.098
	none	0.098 (0.010,0.209)	0.099

*CLR measures the width of the confidence bounds as (upper bound – lower bound / 2).

Table 3.2. Parameter estimates for the ERR/10mGy due to gamma radiation (β_1) and tritium (β_2).

	Order constrained prior	Parameter Estimate					
Leukemia		β_1	90% HPD	CLR*	β_2	90% HPD	CLR*
	$\beta 2 \geq \beta 1$	0.035	-0.032,0.110	0.071	0.299	0.029,0.705	0.338
	none	0.053	-0.025,0.142	0.084	0.141	-0.325,0.645	0.485
Non-CLL Leukemia							
	$\beta 2 \geq \beta 1$	0.083	-0.018,0.206	0.112	0.345	0.050,0.818	0.384
	none	0.176	0.010,0.374	0.182	-0.281	-1.129,0.552	0.841

*CLR measures the width of the confidence bounds as (upper bound – lower bound / 2)

Chapter 4

Tritium: Relative Biological Effectiveness and cancer risk

4.1 Abstract

The relationship between exposure to tritium, a source of beta radiation, and risk of cancer has received little attention in the epidemiological literature. For radiation protection purposes, and for radiation risk assessments, tritium dose and gamma radiation doses are often combined. This is done under the assumption that the biological effectiveness of the absorbed dose of tritium relative to the absorbed dose of gamma radiation, or RBE, is 1. In this paper, we utilize Bayesian methods in order to evaluate the excess relative rate (ERR) of leukemia and leukemia excluding CLL per unit of absorbed dose of tritium as well as its RBE compared to absorbed dose of gamma radiation. The ERR/10mGy and associated 90% highest posterior density (HPD) of leukemia and leukemia excluding CLL are 0.272 (0.024, 0.660) and 0.304 (0.044, 0.753), while the RBEs are 8.36 and 5.36, respectively. This is the first empirical estimate of tritium's relationship to cancer risk based on epidemiological data.

4.2 Introduction

The relationship between intake of tritium and risk of cancer has never been directly assessed in the epidemiological literature [3]. Tritium is a source of beta radiation exposure. Unlike x-rays or gamma radiation, tritium only penetrates a short distance in human tissue. When inhaled or ingested, often in the form of tritiated water (HTO), tritium disperses throughout the human body.

For radiation protection purposes, and for radiation risk assessments, absorbed doses of tritium and gamma radiation are often combined. This is done under the assumption that tritium and gamma radiation doses both have equivalent biological effectiveness. The biological effectiveness of different forms of radiation is often expressed numerically as a weighting factor (w_f) that is based on evidence of the relative biological effectiveness (RBE). RBE is defined as "the ratio of the absorbed dose of a reference radiation to the absorbed dose of the radiation under consideration that is required under similar conditions to produce an identical level of biological response in a particular animal or cellular study." To date, the estimated cancer risk per unit dose of beta radiation from tritium is considered equal to that of gamma radiation. However, there is growing concern that this assumption regarding the RBE for tritium is incorrect.

In a recent systematic review, Little and Lambert suggest that the tritium RBE is more likely to be a value of 2-3, or even as high as 10 [1]. This assessment is based on evidence from experiments in which animals or cells are exposed to chronic, low level gamma or tritium based radiation. However, there is very little direct evidence regarding the carcinogenic effect of tritium in humans. Little and Lambert noted the need for

epidemiological studies of cancer risk following occupational tritium exposures, pointing specifically to the possibility of such research among workers at the US Department of Energy's Savannah River Site.

In this paper we investigate the association between estimated cumulative radiation dose from tritium and leukemia mortality among SRS workers. Examination of the association between tritium and cancer risk is hindered by the fact that exposures to tritium at SRS were usually received chronically at low levels, and occurred simultaneous with occupational exposures to gamma radiation and neutrons. Sparse data also make it difficult to obtain reliable parameter estimates in linear excess relative rate models, which are used to evaluate cancer-radiation dose relationships [28, 63]. However, recent developments in statistical software allow simple implementation of Markov Chain Monte Carlo (MCMC) simulation methods for conducting Bayesian analysis, which provide a way to explore this relationship. Bayesian methods have received increased attention in the epidemiologic literature as a means for researchers to integrate information external to a study into their statistical analyses in order to address previously intractable epidemiologic inquiries [33, 34, 36, 64]. In the case of sparse data, informative priors can be used to improve estimation of an exposure-disease relationship.

The goals of this study are two-fold. First, we will utilize Bayesian methods to examine the relationship between tritium and cancer risk among the Savannah River Site (SRS) occupational cohort. The SRS cohort is unique in that it has the expressed purpose of producing tritium for use in nuclear weapons. Thus, SRS has maintained detailed records of individual radiation dose received from the beginning of operations in 1950 to present. This includes disaggregated values for absorbed tritium and gamma radiation

dose [65]. A recent systematic review highlighted SRS as one of the few facilities which has the potential to inform the relationship between tritium dose and cancer risk [3]. Second, since the SRS cohort data include individual gamma dose, we will also be able to provide an estimate of the RBE of tritium relative to gamma radiation in the context of an observational epidemiologic study.

4.3 Materials and Methods

4.3.1 Savannah River Site cohort

The Savannah River site (SRS) is a nuclear fuel facility near Aiken, SC. This facility is unique among US facilities in that it is the only location with the expressed purpose of producing tritium, a radioisotope of hydrogen that emits beta radiation via the release of electron energy as it decays from H^3 to He^3 [5]. Activities began in 1951, with the first tritium production reactor going critical in December 1953. SRS consists of numerous "areas" defined by health physicists based on the activities that take place at each. These areas include: nuclear reactors (100 Area), separations for processing irradiated materials (200-F and H Areas), fuel and target fabrication (300-M Area), heavy water extraction (400 Area), Laboratories (700 Area), and administration and non-nuclear facilities to support plant activities. E.I du Pont de Nemours and Company led operations at the site until March 31, 1989 when Westinghouse Savannah River Company took over, who operate the facility at present [66, 67].

This cohort includes 21,204 employees who were hired by du Pont between calendar years 1950-1987, and who worked at least 90 days with no history of

employment at another DOE facility. In addition, workers were excluded if they were missing information on gender, date of birth, name, SSN, or date of first hire. This leaves a cohort of 18,883 workers for whom individual, annual dose records have either been computed or estimated via previous research. Since most of the cohort was hired before 1960, some workers may have up to forty years of dose records. Workers were followed through 2002 to obtain vital status information. In the current analysis, the outcomes, leukemia and leukemia excluding CLL are based on International classification of disease in the United States (ICD) codes. Through 1990, ICD-8 codes were used, while ICD-9 codes were used thereafter. The identification of leukemia and leukemia excluding CLL in ICD 8 and 9 were the same. For all leukemia, the codes are 204-207. Those with code 2041 represent cases of CLL. One reason that we examine leukemia in this analysis is because it is less subject to confounding by smoking than outcomes such as lung cancer. Analyses excluding CLL are conducted because of the longer time course of disease of CLL compared to other forms of leukemia.

Annual whole body dose (WBD) represents the sum of deep tissue gamma dose, tritium dose, and neutron dose converted to biologically equivalent gamma dose with a weighting factor, w_f. Estimated annual WBD values are available for all employment years. 6% of these records were estimated using a 'nearby' method described in Richardson et al [67]. In total, whole body dose records sum to 520.8 person-Sieverts (Sv) of collective dose. Annual whole body dose estimates were utilized in order to derive annual tritium dose estimates for those person-years with missing information regarding the contribution of tritium to their cumulative annual whole body dose. In total, there are 56.2 Sv of individual, annual tritium dose records, of which 4.3 Sv (7.7%) were

estimated through use of a job-exposure matrix described by Hamra et al [65]. Neutron doses are quantified and integrated into whole body dose using a w_f that changes based on the amount of absorbed dose received by the worker. As a result, absorbed neutron doses are unavailable for this analysis, and the reweighted, absorbed doses that are received are treated as equivalent to gamma radiation dose. Thus, neutron and other internal doses are not considered in this analysis.

Since we are using disaggregated gamma and tritium dose, and calculating values of the RBE, the relevant value of individual dose is 'absorbed,' which is expressed in Grays (Gy). This is distinct from more recent radiation epidemiology that has focused interest on 'effective dose,' expressed in Sieverts (Sv). The difference between these two measures is that doses combined into effective dose takes account of the w_f for more highly effective forms of radiation. We are able to obtain absorbed dose because the aggregated whole body dose utilizes a w_f of 1. Thus, absorbed gamma and tritium doses are simply the two components of whole body dose.

4.3.2 Statistical model

To examine the relationship between tritium dose and cancer, as well as its relative biological effectiveness (RBE) compared to gamma radiation, we will utilize the data from the SRS cohort as previously examined by Richardson and Wing [28]. This study examined the dose-response relationship of individual, cumulative whole-body dose radiation and leukemia, including and excluding CLL.

Dose-response relationships are estimated using an excess relative rate (ERR) model of the form:

$$RR = e^{\alpha i}(1+\beta_1 g + \beta_2 t)$$

where '$e^{\alpha i}$' indexes the baseline rate within strata 'i', β represents the ERR for a given exposure, 'g' represents cumulative gamma radiation dose, and 't' represents cumulative tritium dose.

As in prior analyses we utilize a nested case control design by creating risk sets based on a set of selected matching factors. Eligibility criteria for inclusion of individuals in this study, as well as the choice of matching factors for conducting their study are the basis for this analysis, and are clearly documented in their work. The distributions of leukemia and leukemia excluding CLL cases with respect to matching factors (including attained age, sex, race, pay code, birth cohort, and employment status) are presented in Richardson and Wing [28]. In total, we observe 84 cases for leukemia, 22 of which were CLL cases. The average risk set includes 480 controls. The smallest risk set includes four controls.

The distinctions between the current work and that of Richardson and Wing are the use of disaggregated radiation dose and Bayesian methods. The annual records for gamma radiation and tritium dose are summed for each year from the beginning of employment until the end of observation to obtain cumulative gamma and beta dose.

We utilize a Bayesian modeling approach in order to integrate prior information into our analysis, described below. In order to inform estimation of the ERR associated with tritium and gamma radiation, we incorporate two pieces of information based on what is known about the etiology of radiation exposure and cancer risk. Bayesian inference provides a way to integrate prior knowledge into an epidemiologic analysis in

order to provide more accurate and stable parameter estimation. In addition to providing a model for relationships between exposures and outcomes we can incorporate knowledge about the dose-response relationships based on information including evidence from previous epidemiological or radiobiological studies.

First, we incorporate our prior belief that gamma radiation exposure has either no relationship to cancer risk or increases the risk of cancer (i.e., exposure to gamma radiation does not diminish a worker's risk of cancer). This substantive prior is based on evidence that gamma radiation is an established mutagen and carcinogen [68]. A simple way to build this knowledge into our model is to truncate the prior distribution of β_1 at a lower limit of 0. In addition, tritium emits low-LET β^- radiation, and is an established mutagen and carcinogen [8, 10, 69]. Thus, we specify the same prior for the lower bound of the ERR for tritium by specifying $\beta_2 \geq 0$.

Second, we incorporate our prior knowledge that ionizing radiation from tritium intakes is at least as effective as gamma radiation at causing cancer (i.e., the RBE for tritium is 1 or greater). This prior belief is based upon evidence from the in vivo and in vitro literature, which suggests that, for a given amount of absorbed dose, tritium is at least as biologically effective as gamma radiation at causing a variety of biological endpoints of potential relevance to cancer induction including chromosomal aberrations and induction of cancer in experimental animals [1, 18, 19]. We integrate this knowledge into our analysis by specifying an order constrained prior such that $\beta_2 \geq \beta_1$ [62]. In an MCMC implementation of the analysis, this translates to drawing samples of β_2 that are never less than β_1. Examples of using Bayesian analysis for this form of restricted

parameter estimation have been shown elsewhere in the literature [62, 70]. The form of the distribution for the parameters is specified as normal but non-informative, where $\beta_1 \sim N(\mu=0, \sigma^2=100{,}000)$ and $\beta_2 \sim N(\mu=\beta_1, \sigma^2=100{,}000)$. In other words, the form of the posterior distribution is not informed by this specification.

In order to conduct this analysis in a Bayesian framework, we utilize the SAS procedure MCMC (V 9.2 Cary, NC). An estimate of the RBE is obtained at each step in the MCMC process, and is calculated as β_2/β_1. This simulation based approach allows us to obtain uncertainty bounds around the estimate of the RBE. Posterior distributions are presented as 90% highest posterior density (HPD) intervals for consistency with radiation epidemiology literature.

Parameter estimates may be sensitive to the choice of prior information included in a model, especially when data are sparse [54]. Sensitivity analyses were conducted under alternative priors for the parameters (β_1 and β_2). All models were run 3 times for 100k iterations with a 10k burn-in and 3 distinct starting values to emulate a Gelman-Rubin diagnostic check to ensure that results are not sensitive to starting values of the β_1 and β_2. Final models were run for 10 million iterations with a burn-in of 100,000 iterations to minimize simulation error.

4.4 Results

The Appendix includes diagnostic plots for each model in our analysis to confirm convergence of the Markov chain. Table 4.1 summarizes the distribution of cases by categories of cumulative whole body dose, gamma dose, and tritium dose. This table

shows most cases receiving less tritium than gamma dose over the course of employment. Figure 4.1 plots the relationship between cumulative gamma dose and tritium dose among leukemia cases (n=84). The Pearson correlation coefficient is 0.79 for cases. When examining all cases and controls, the Pearson correlation coefficient for the relationship between gamma and tritium dose is 0.34.

Table 4.2 presents the distribution of gamma and tritium dose of cases and controls for leukemia and leukemia excluding CLL. Cumulative dose is lagged 3-years from the date of cancer mortality. The distributions of dose for cases and controls are similar for each outcome. Dose distributions are highly skewed to the right, with most cases and controls accruing low cumulative dose over time employed at SRS. For the most part, employees receive less cumulative tritium dose than gamma dose. However, there are rare instances of high beta radiation exposures. With regard to leukemia, the maximum gamma dose received by controls is 469.7 mGy as compared to the maximum beta dose of 155.68 mGy for a sampled control.

Table 4.3 summarizes the ERR/10mGy dose received for leukemia and leukemia excluding CLL in our primary analysis. The ERR/10mGy (90% HPD) of leukemia associated with tritium and gamma radiation are 0.282 (0.027, 0.678) and 0.044 (0.000, 0.108), respectively. This yields an RBE of 6.38 (1.00, 36.09). With regard to leukemia excluding CLL, the ERR associated with tritium and gamma radiation are 0.338 (0.048, 0.805) and 0.087 (0.000, 0.195), respectively. This yields an RBE of 3.88 (1.00, 16.80).

4.5 Sensitivity analysis

Table 4.4 summarizes the results of a sensitivity analysis which tests the change in ERR/10mGy and RBE estimates upon removal of prior information from the model. The full model (model 1), which is the subject of our primary analysis summarized in table 4.3, contains two levels of prior information discussed above:

a) Specification of $\beta_2 \geq \beta_1$.

b) Specification of $\beta \geq 0$.

Each subsequent model removes a level of prior information until the final model where the prior specification of β_1 andβ_2 is normal but non-informative. Model 2 excludes the prior specification of $\beta_2 \geq \beta_1$. Model 3 excludes any restrictions on parameter estimation and utilizes a non-informative prior in the form of a normal distribution where β_1 andβ_2 are N(μ=0, σ^2=100,000).

The estimates of β_1 and β_2 for both outcomes, leukemia and leukemia excluding CLL, respond similarly to integration of prior knowledge (table 4.4). When the model does not include any prior knowledge, β_1 and β_2 include negative values in the 90% HPD. Specifying β_1 and $\beta_2 \geq 0$ in model 2 improves precision of both model parameters. Addition of $\beta_2 \geq \beta_1$ also improves estimation of the parameters by reducing the imprecision of each..

Estimates of the ERR/10 mGy of leukemia and leukemia excluding CLL due to tritium are similar, 0.282 (0.027, 0.678) and 0.338 (0.048, 0.805), respectively. These estimates exhibit modest change from model 1 compared to model 2. However, when the

model includes no prior information, estimation of the ERR/10 mGy for leukemia and leukemia excluding CLL are 0.141 (90% HPD: -0.325, 0.645) and -0.281 (90% HPD: -1.129, 0.552), respectively.

RBE estimation is most precise in model 2. For both leukemia and leukemia excluding CLL, estimates of the upper level of the 90% HPD in model 1 are larger than model 2. However, the distribution of the RBE in model 2 includes values less than 1, which are not supported by the experimental evidence summarized by Little and Lambert [1]. The median of the posterior distribution for the RBE for models 1 and 2 are above the null value. Compared to our full model in table 4.3, estimates of the RBE with regard to leukemia for models 2 and 3 are 4.79 and 2.64, respectively. With respect to non-CLL leukemia, the RBE estimate obtained in model 2 is 2.25. We do not report an RBE for model 3 when non-CLL leukemia is the outcome because the value is negative, which is not a plausible estimate for a ratio of two linear slope parameters.

4.6 Discussion

The goals of this analysis are two-fold. First, we provide an empirical estimate of tritium's relationship to cancer risk based on observational data. Second, using this information, we obtain an estimate of the RBE of tritium relative to gamma radiation, which further supports the existing evidence from in vitro and in vivo studies that this value is above the null value of 1 [1].

Evidence regarding the relationship between tritium and cancer risk has only been available from in vivo and in vitro studies. While animal models may examine cancer incidence and mortality, cellular evidence is limited to different forms of cellular and

chromosomal damage, such as acentric rings. In both cases, extrapolating this knowledge to inform human cancer risk estimation is a challenge for epidemiologists. First, we do not assume exchangeability of risk estimates from cancer precursors to cancer events, or cancer outcomes across species. Second, there is no established metric with which we can convert results of animal and cellular data to apply in human cancer risk models.

Bayesian inference provides an intuitive means for incorporating animal and cellular knowledge into observational epidemiology studies. By incorporating this knowledge into a Bayesian model as prior knowledge, we are able to test its ability to inform estimation of the posterior distribution of parameter estimates of interest.

Truncation of the parameter estimates at a lower bound of 0 leads to estimates of the RBE greater than 1. Although estimates of the RBE are increased as additional information is integrated into the model, they do not exceed the plausible range of values suggested in the radiobiological literature. Little and Lambert estimate an RBE of 2.49 based on experimental evidence of chronic, low dose exposure scenarios similar to what would be expected in an occupational workplace. Our results suggest that the biological effectiveness of tritium relative to gamma radiation may be greater than previously suggested by animal and cellular evidence.

Our analysis provides an example of where Bayesian methods provide a clear benefit over frequentist analysis. Sparse data results in poor estimation of the relationship between tritium dose and risk of leukemia excluding CLL. By simply constraining estimation of β_2 and β_1 based on the knowledge that radiation is not protective of cancer risk, we are able to obtain stable estimates of the ERR of leukemia and leukemia

excluding CLL associated with tritium dose. We note a major improvement in the 90% HPD when we truncate values at a lower bound of 0. The change in the 90% HPD is much less pronounced upon adding the constraint that $\beta_2 \geq \beta_1$.

Our analysis also faces some limitations. For instance, we are unable to examine absorbed neutron dose independent of absorbed gamma dose, due to the fact that all values are converted 'effective' dose values to account for the greater biological effectiveness of neutrons. In addition, the bulk of occupations and areas of employment where tritium dose was prominent were dominated by males, with few females obtaining non-negligible tritium dose, or the outcome of interest. Interpretation of our results should consider these limitations.

Despite the fact that we obtain stable parameter estimates by integrating external knowledge, our analysis may benefit from integration of data from other nuclear facilities. Little and Wakeford discuss facilities in Canada and the UK where information on tritium dose is recorded, but has not been examined [3]. It would be possible to combine this data with our current analysis to improve estimation of tritium's relationship to cancer and RBE referenced against gamma radiation.

In addition, it would be informative to examine the relationship between tritium dose and other cancers, such as all-cancer. In order to do so we will need to account for bias due to smoking, which has not been possible in the SRS cohort with the current data. However, we hope to do this by use of bias modeling, or by examining other occupational cohorts that have smoking history data [71].

Finally, the NIOSH office of compensation analysis and support has the expressed goal of assisting in compensation for work related illness, specifically cancer, among nuclear facility employees via determining probability of causation and conducting dose reconstruction for workers missing personal dose measures. Aggregated whole body dose is used to determine the support of these claims. To this end, it is necessary to better understand the RBE of tritium to evaluate its possible influence on work-related cancer illness.

Table 4.1 Distribution of cases by categories of individual cumulative whole body dose, gamma, and tritium dose.

Outcome		Dose (in mGy)								
	Radiation Form									
Leukemia		0	< 0 - <5	5 -<10	10-<20	20 - <40	40 - <80	80 - <160	160- <320	≤ 320
	Tritium	30	34	9	3	5	3	0	0	0
	Gamma	5	27	8	9	10	7	13	4	1
	Whole body dose	5	26	9	8	8	9	13	4	2
Leukemia excluding CLL										
	Tritium	24	25	4	2	4	3	0	0	0
	Gamma	4	20	6	6	4	6	11	4	1
	Whole body dose	4	19	7	6	4	5	11	4	2

Table 4.2. Distribution of radiation dose estimates for cases and controls by categories of leukemia outcome.

	Dose Distribution (in mGy)			
	Gamma		Tritium	
	mean (SD)	median, max*	mean (SD)	median, max*
Leukemia				
cases	41.64 (63.86)	10.93, 320.3	5.47 (11.06)	0.89, 49.97
controls	38.05 (64.96)	7.13, 469.7	5.18 (10.62)	0.55, 155.7
Leukemia (non-CLL)				
cases	47.65 (70.24)	10.38, 320.0	5.97 (12.40)	0.55, 49.97
controls	38.73 (65.43)	7.30, 467.5	5.33 (10.76)	0.55, 155.7

*All dose distributions are characterized by a minimum value of 0 mGy.

Table 4.3. Estimated associations between radiation doses from gamma rays and tritium intakes and leukemia mortality (ERR/10mGy) among Savannah River Site workers, and estimated relative biological effectiveness (RBE) of tritium.*

	Parameter Estimate					
Outcome	β_1	90% HPD	β_2	90% HPD	RBE	90% HPD
Leukemia	0.044	0.000,0.108	0.282	0.027,0.678	6.38	1.00,36.09
Non-CLL Leukemia	0.087	0.000,0.195	0.338	0.048,0.805	3.88	1.00,16.80

*primary analysis indicates that all prior information discussed in 'materials and methods' are included in this MCMC model.

Table 4.4. Sensitivity Analysis comparing results of primary analysis to models with levels of prior information removed.

Model	Prior included*	Parameter Estimate					
Leukemia		β_1	90% HPD	β_2	90% HPD	RBE	90% HPD
1	Full model	0.044	0.000,0.108	0.282	0.027,0.678	6.38	1.00,36.09
2	$\beta \geq 0$	0.049	0.000,0.118	0.236	0.000,0.630	4.79	0.00,30.68
3	none	0.053	-0.025,0.142	0.141	-0.325,0.645	2.64	-29.56,36.81
Non-CLL Leukemia							
1	Full model	0.087	0.000,0.195	0.338	0.048,0.805	3.88	1.00,16.80
2	$\beta \geq 0$	0.103	0.000,0.226	0.231	0.000,0.687	2.25	0.00,12.44
3	none	0.176	0.010,0.374	-0.281	-1.129,0.552	n/a	n/a

*Prior information is described in the Sensitivity Analysis section.

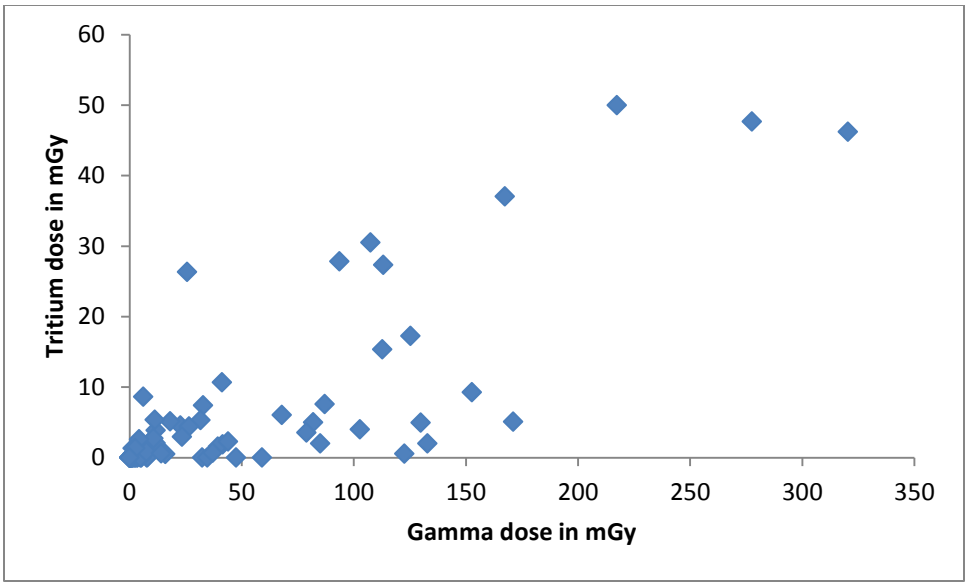


Figure 4.1 Scatter plot of cumulative tritium dose by cumulative gamma radiation dose among leukemia cases. Pearson correlation coefficient = 0.79.

Chapter 5

Conclusion

5.1 Overview

Bayesian methods have gained popularity in epidemiology as a means for handling statistical analyses that prove problematic when using standard maximum likelihood estimation techniques. Using MCMC simulation, the machinery of Bayesian analysis, epidemiologists can handle a variety of problems, such as highly correlated data, multiple comparisons, and bias due to sparse data [33, 34, 36, 72]. We have provided an example of where Bayesian methods provide a clear benefit over standard regression modeling. First, we demonstrate that Bayesian methods provide a simple method for integrating knowledge from experimental research into an occupational cohort study via an order constrained prior. Second, we used this tool in order to provide the first epidemiologically based estimates of the excess relative rate of leukemia and leukemia excluding CLL associated with exposure to beta radiation emitting tritium. The use of a benefit of an order constrained prior is that it does not require arbitrary parameter specification due to lack of an established metric for converting animal and cellular data into informative priors for human data research. Instead, we simply specify the direction

of effect for a parameter relative to another parameter in the same model. We believe this will be a useful tool for epidemiologists in future research.

5.2 Bayesian methods for epidemiologists

Frequentist statistical methods have dominated the field of epidemiology for decades. Despite the fact that Bayesian methods have existed since the 18th century, epidemiologists and statisticians alike have found implementation difficult, due largely to demanding statistical procedures and very specialized software implementation, such as WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/). However, recent additions to common statistical software packages (such as SAS) and a resurgence of interest in Bayes' theory have led to increased interest in and utilization of Bayesian methods [34].

In chapter 2, we detail the process through which MCMC simulation obtains estimates of the posterior distribution of parameters given the data. We focus attention on the fact that when a researcher has no prior knowledge to integrate into an MCMC procedure, the estimates of the parameters of interest will be the same as those obtained from standard MLE analysis. We provide this information to motivate use of Bayesian methods by frequentist statisticians who may be otherwise reluctant to utilize novel statistical analytic techniques.

5.3 Order constrained priors and truncation

The use of an order constrained prior has its roots in modeling of dose response relationships when the direction of the relationship is known [62]. Through integrating the simple restriction that increases in the unit of some exposure will lead to a monotonic

increase or decrease in the observed outcome (whether it be an odds, rate, or other measure), a researcher is able to improve the precision of parameter estimation. We extend this simple approach to a scenario where two distinct exposures are measured, and integrate the knowledge that the magnitude of one exposure is greater than the other. Regarding leukemia excluding CLL, we see a negative estimate of the ERR/10mGy associated with beta radiation exposure. The negative relationship represents a change in direction compared to the same model where the outcome is leukemia including CLL. This would not be expected, since the two outcomes are related. We might expect a modest bias away from the null by including CLL with other forms of leukemia, but not a shift in the direction of effect. In addition, a reduction in the risk of cancer due to radiation exposure is not supported by any existing research. Radiation is known to have either no association or a positive association with cancer. By integration of an order constrained prior that provides a direction for estimation of the association of beta radiation with cancer risk, we are able to obtain stable and more plausible parameter estimates.

This work introduces a simple method for integrating knowledge from experimental studies into observational research. Typically, if a researcher wants to integrate prior information from an experimental study into epidemiologic research, it is necessary to alter the values obtained due to a number of factors. When comparing animal models to human research, estimates are not exchangeable because hazardous agents may exhibit differential effects across species. When considering cellular level studies, outcomes are not the same as those that are of interest in observational research. While epidemiological research will focus on cancer incidence or mortality, cellular

studies are more often interested in cancer precursors, such as chromosomal aberrations and DNA damage. These results may suggest a relationship between a hazardous agent and cancer, but may not directly inform it. Finally, error distributions around parameter estimates of interest are often overconfident, since the number of observed outcomes and the exact amount of exposure is under the control of the investigator. Through the use of truncation in a Bayesian analysis, we are able to integrate knowledge of the direction and magnitude of effect of tritium without the need to explicate a prior distribution for the parameter.

5.4 Tritium risk and relative biological effectiveness

The results from our study using a data-driven, order constrained prior illustrate a means of integrating knowledge from animal and cellular studies into observational research. When we specify a non-informative prior, the estimated ERR/10mGy due to tritium (β_2) for leukemia and leukemia excluding CLL are 0.141 (90% HPD: -0.323, 0.649) and -0.281 (90% HPD: -1.136, 0.548), respectively. When we truncate $\beta_2 \geq \beta_1$, the estimated ERR/10mGy due to tritium for leukemia and leukemia excluding CLL are 0.298 (90% HPD: 0.027, 0.702) and 0.344 (90% HPD: 0.049, 0.817), respectively.

Our results regarding the ERR of leukemia and leukemia excluding CLL per unit dose of tritium based beta radiation are the first of their kind in the epidemiological literature. By modeling tritium and gamma radiation in a single ERR model we are also able to obtain estimates of the RBE for tritium relative to gamma radiation by simply calculating the ratio of samples drawn at each iteration of the MCMC process. These models include specification of prior knowledge based on radiation epidemiology

literature, knowledge of occupational protection standards, and the experimental literature regarding the RBE for tritium.

Even when we exclude information regarding the RBE of tritium, the results of our analysis are consistent with the experimental evidence. In models that only include truncation of the ERR per unit dose for beta and gamma radiation, the RBE estimates for leukemia and leukemia excluding CLL are 4.79 and 2.25, respectively. This is similar to the estimate of 2.49 recommended by Little and Lambert for a chronic, low dose exposure to beta and gamma radiation [1]. In addition, in a model with no informative prior, the RBE for leukemia as the outcome of interest is 2.25.

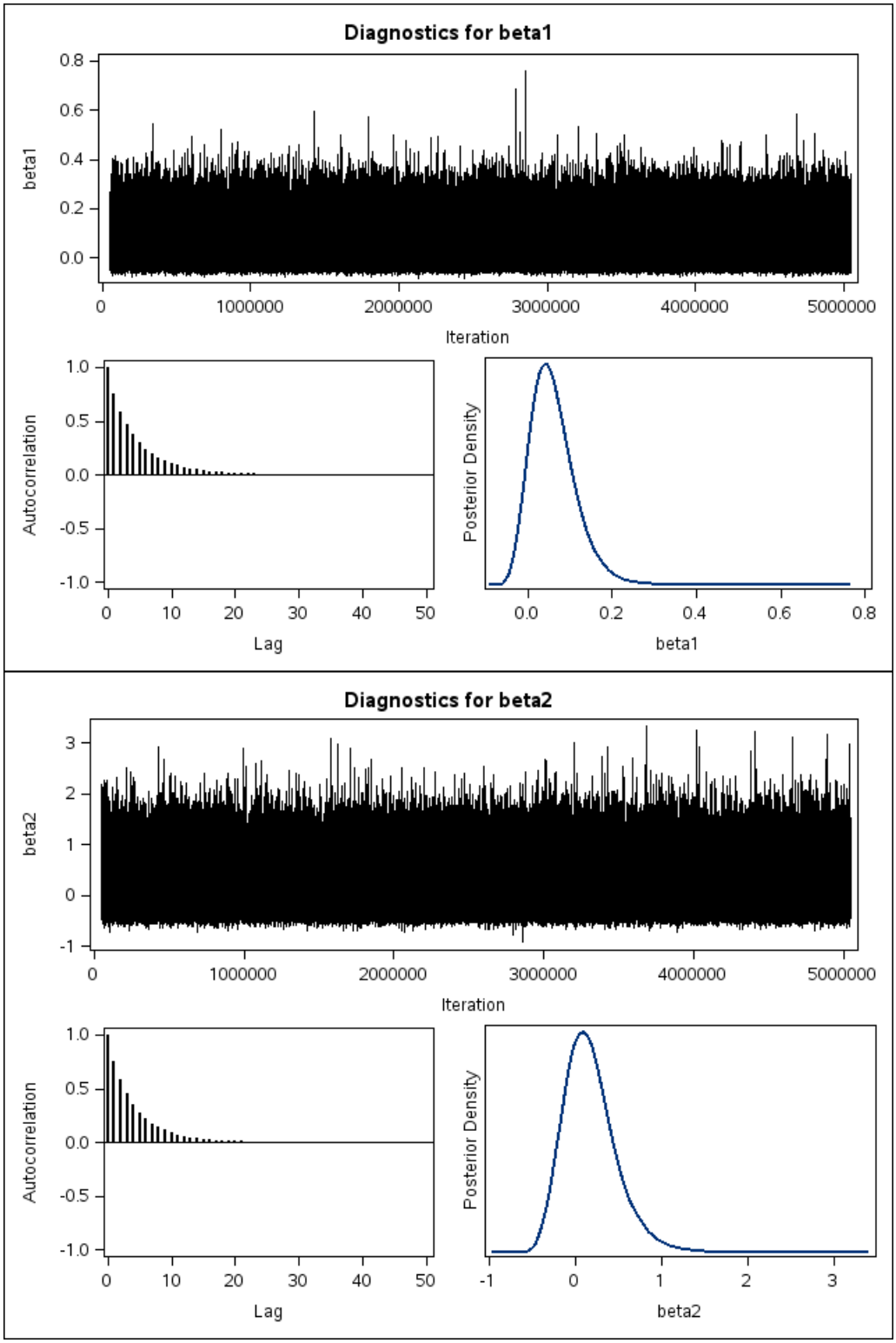
There are a number of in vitro and in vivo studies examining the adverse health effects of tritium exposure. The outcomes range from chromosomal damage and cell death to cancer outcomes such as leukemia. However, our work is the first example of a research team examining the relationship between tritium dose and cancer risk in an observational cohort study.

5.5 Summary and Significance

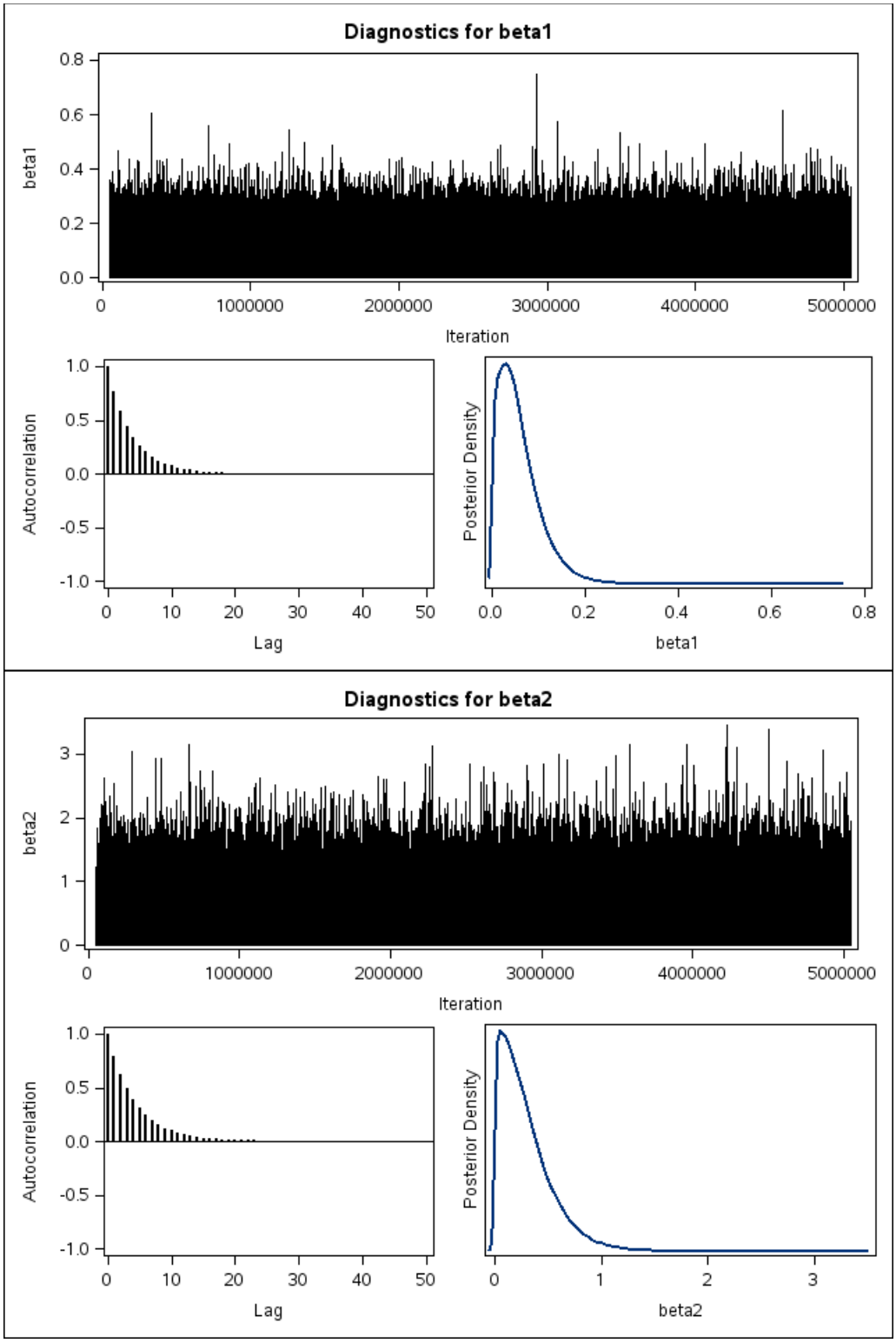
Through the use of Bayesian methods, we obtain the first empirical estimates of the relationship between tritium exposure and cancer risk in an observational study. In addition, we are able to confirm the findings of experimental studies, which suggest that tritium based beta radiation is more biologically effective than gamma radiation. Finally, through our analysis, we are able to demonstrate an intuitive and simple approach to combining evidence from in vitro and in vivo studies into analysis of an observational epidemiology dataset.

The NIOSH office of compensation analysis and support has the goal of assisting in compensation for work related illness, specifically cancer, among nuclear facility employees via determining probability of causation and conducting dose reconstruction for workers missing personal dose measures. To this end, it is necessary to provide proper estimation of the RBE of tritium to evaluate its possible influence on work-related illness. This information may improve the quality of dose reconstruction and risk estimation for nuclear workers. Additionally, exposure to tritium is not exclusive to the occupational setting. The Exelon energy facility of Illinois was recently sued by community members due to evidence of discarding tritium waste into local waterways. Thus, an estimate regarding the relationship between tritium exposure and cancer may be helpful to those who receive environmental exposures.

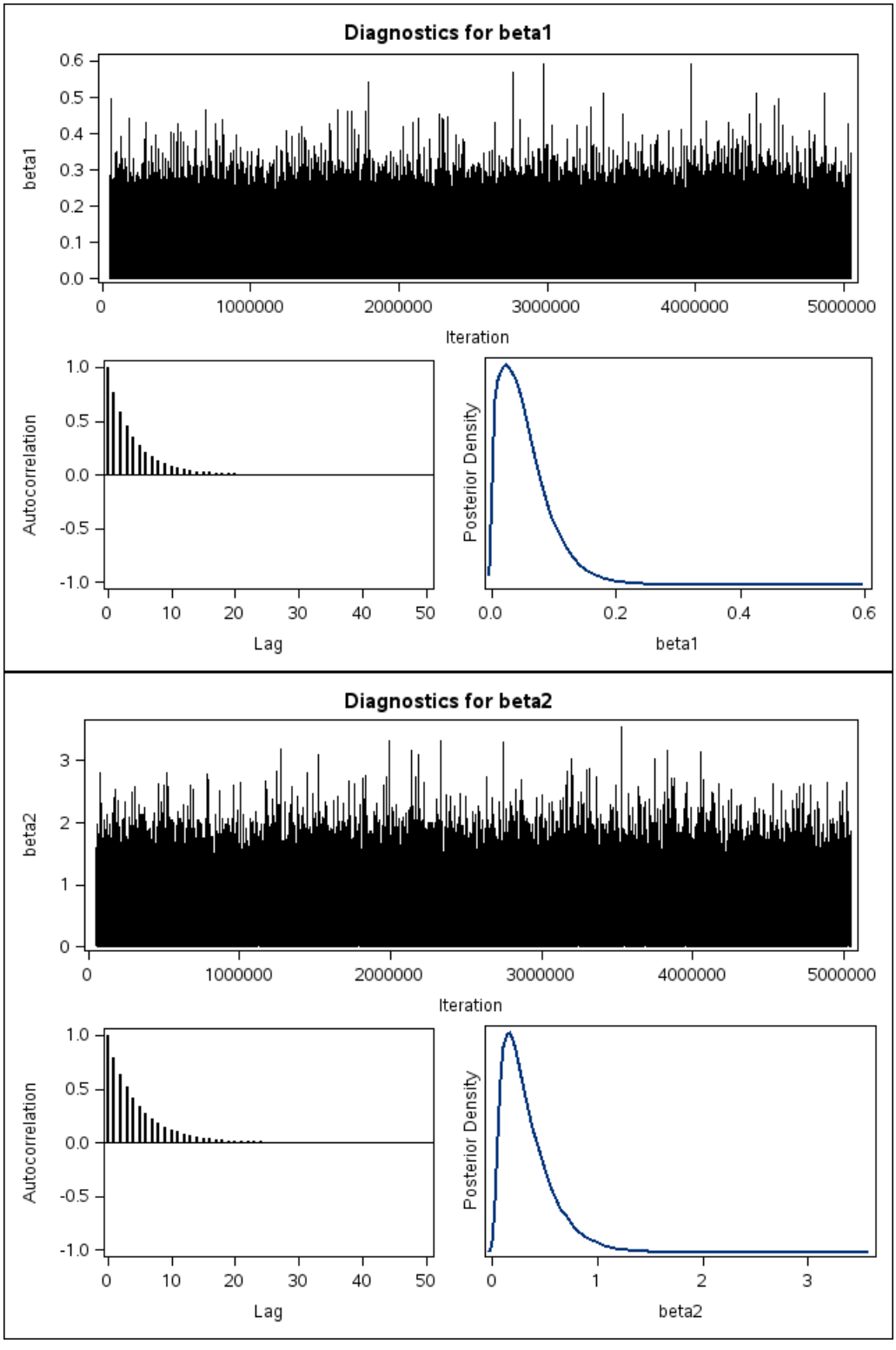
APPENDIX



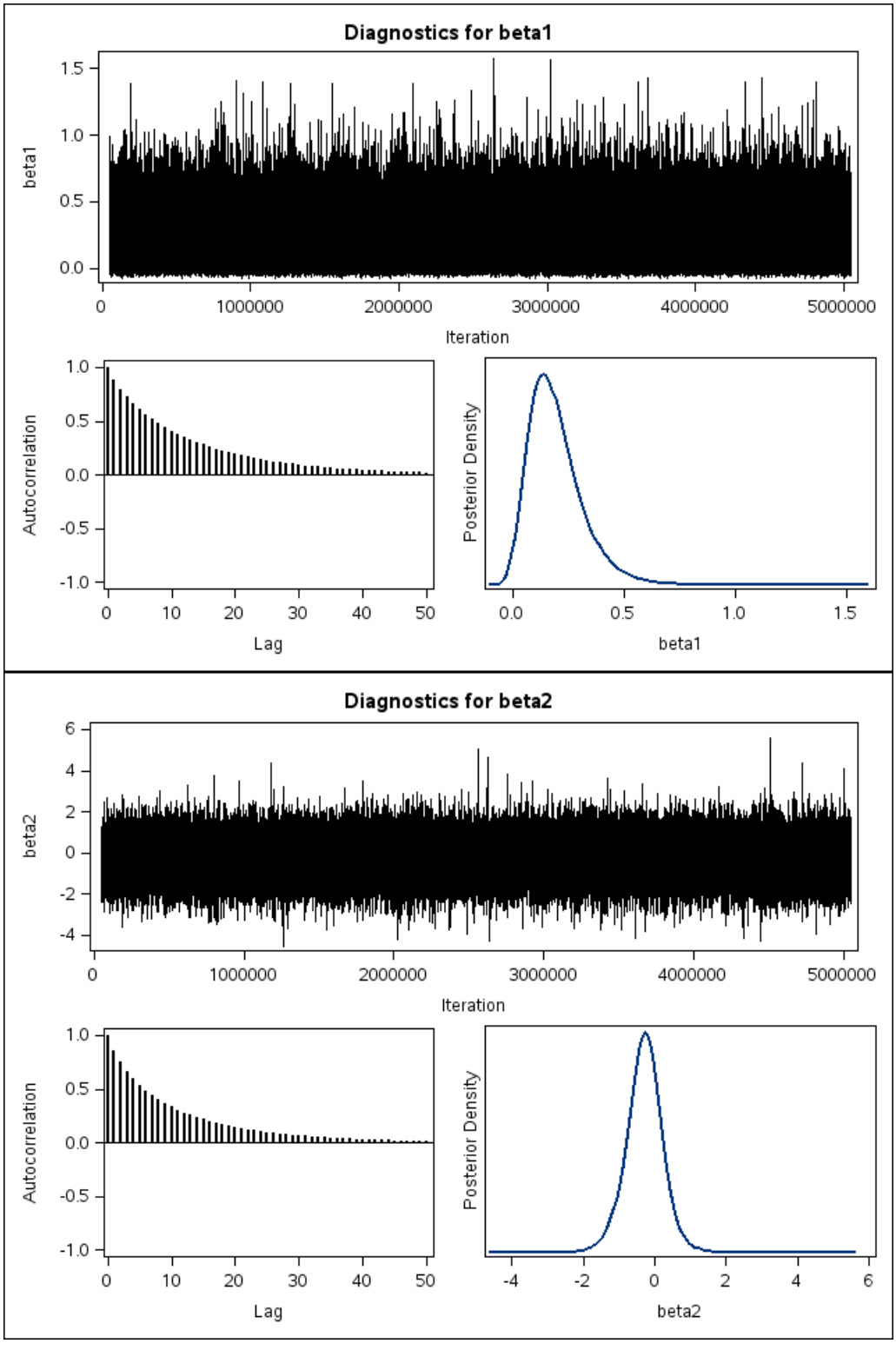
Chapter 4, null model diagnostic plots for β_2 and β_1 examining ERR/10 mGy for leukemia



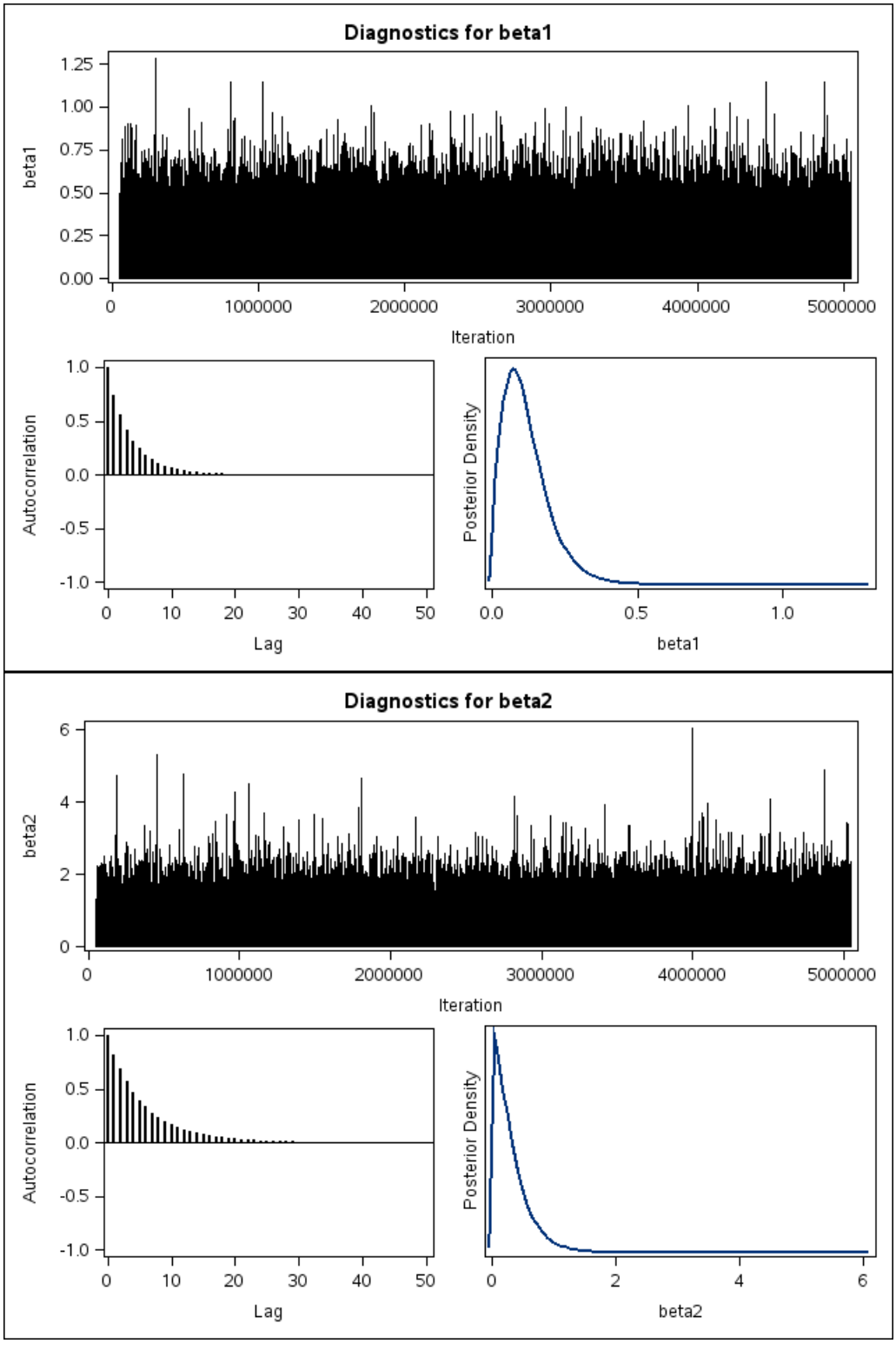
Chapter 4, diagnostic plots for a model with specification of $\beta_1 \geq 0$ and $\beta_2 \geq 0$ examining ERR/10 mGy for leukemia



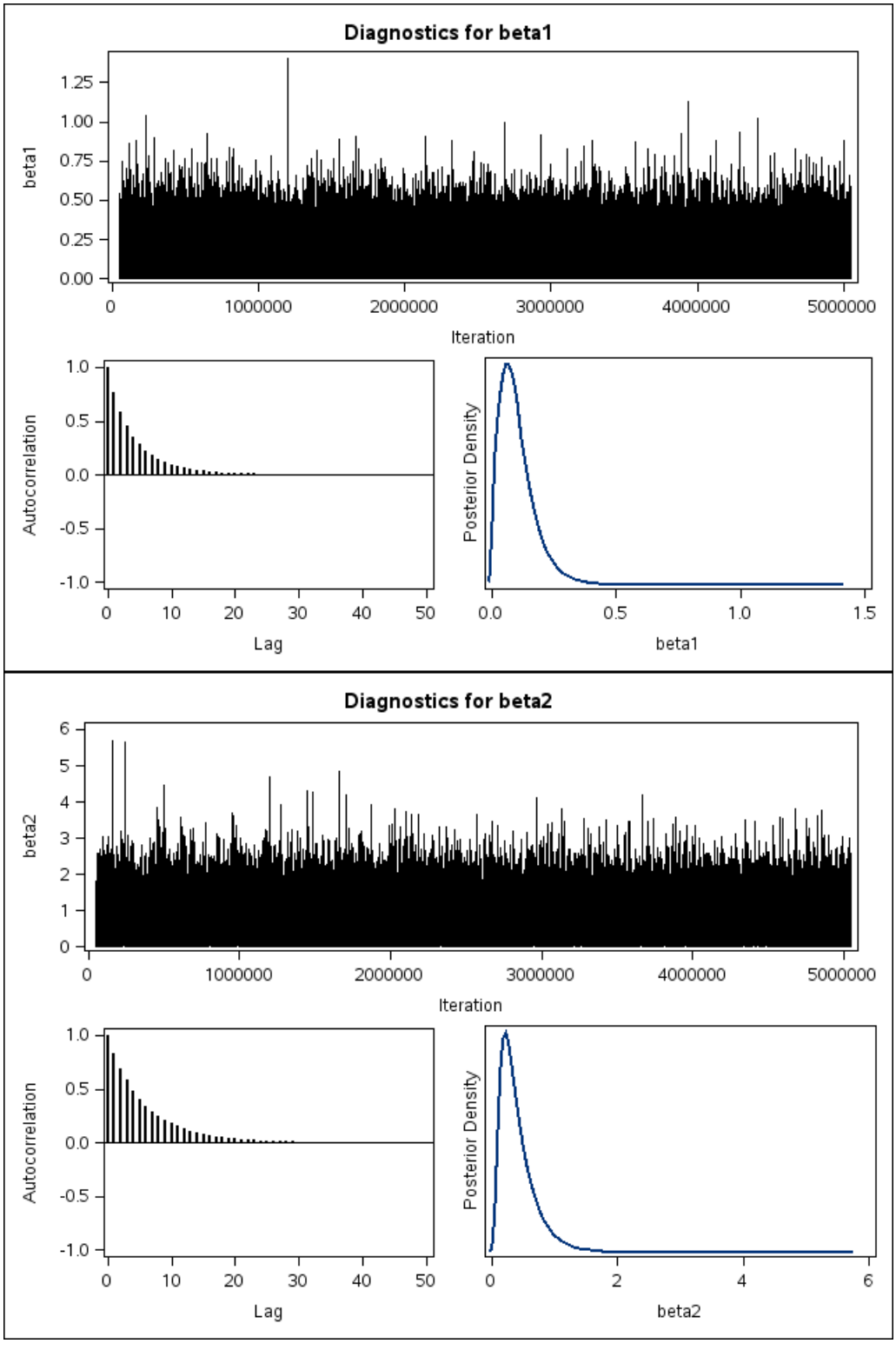
Chapter 4, diagnostic plots for a model with specification of $\beta_1 \geq 0$ and $\beta_2 \geq \beta_1$ examining ERR/10 mGy for leukemia



Chapter 4, null model diagnostic plots for β_2 and β_1 examining ERR/10 mGy for leukemia excluding CLL.



Chapter 4, diagnostic plots for a model with specification of $\beta_1 \geq 0$ and $\beta_2 \geq 0$ examining ERR/10 mGy for leukemia excluding CLL



Chapter 4, diagnostic plots for a model with specification of $\beta_1 \geq 0$ and $\beta_2 \geq \beta_1$ examining ERR/10 mGy for leukemia excluding CLL.

References

1. Little, M.P. and B.E. Lambert, *Systematic review of experimental studies on the relative biological effectiveness of tritium.* Radiat Environ Biophys, 2008. **47**(1): p. 71-93.

2. Hill, R.L. and J.R. Johnson, *Metabolism and dosimetry of tritium.* Health Phys, 1993. **65**(6): p. 628-47.

3. Little, M.P. and R. Wakeford, *Systematic review of epidemiological studies of exposure to tritium.* J Radiol Prot, 2008. **28**(1): p. 9-32.

4. Taylor GA, C., KW, La Bone RD, Wilkie WH, *A History of Personnel Radiation Dosimetry at The Savannah River Site.* 1995.

5. Cragle, D.L., et al., *Mortality among workers at a nuclear fuels production facility.* Am J Ind Med, 1988. **14**(4): p. 379-401.

6. Okada, S. and N. Momoshima, *Overview of tritium: characteristics, sources, and problems.* Health Phys, 1993. **65**(6): p. 595-609.

7. Elliott, A., *Tritium dosimetry.* Nucl Med Commun, 2005. **26**(6): p. 481-2.

8. Potter, C.A., *Application of the ICRP clarification of the tritium metabolic model.* Health Phys, 2004. **87**(4): p. 375-81.

9. Joksic, G. and V. Spasojevic-Tisma, *Chromosome analysis of lymphocytes from radiation workers in tritium-applying industry.* Int Arch Occup Environ Health, 1998. **71**(3): p. 213-20.

10. Straume, T., *Tritium risk assessment.* Health Phys, 1993. **65**(6): p. 673-82.

11. National Research Council (U.S.). Committee to Assess Health Risks from Exposure to Low Level of Ionizing Radiation., *Health risks from exposure to low levels of ionizing radiation : BEIR VII Phase 2*2006, Washington, D.C.: National Academies Press. xvi, 406 p.

12. Goodhead, D.T., *The relevance of dose for low-energy beta emitters.* J Radiol Prot, 2009. **29**(3): p. 321-33.

13. Goodhead, D.T., *Initial events in the cellular effects of ionizing radiations: clustered damage in DNA.* Int J Radiat Biol, 1994. **65**(1): p. 7-17.

14. Nikjoo, H. and D.T. Goodhead, *Track structure analysis illustrating the prominent role of low-energy electrons in radiobiological effects of low-LET radiations.* Phys Med Biol, 1991. **36**(2): p. 229-38.

15. Fairlie, I., *RBE and w(R) values of Auger emitters and low-range beta emitters with particular reference to tritium.* J Radiol Prot, 2007. **27**(2): p. 157-68.

16. Nikjoo, H. and L. Lindborg, *RBE of low energy electrons and photons.* Phys Med Biol, 2010. **55**(10): p. R65-109.

17. Protection, I.C.o.R., *1990 Recommendations of the ICRP*. Annals of the ICRP. Vol. Publication 60. 1991. 1-201.

18. Ujeno, Y., *Relative biological effectiveness (RBE) of tritium beta rays in relation to dose rate.* Health Phys, 1983. **45**(3): p. 789-91.

19. Paquet, F. and H. Metivier, *Are the risks from tritium exposures being underestimated?* J Radiol Prot, 2009. **29**(2): p. 175-81.

20. Edwards, A.A. and D.C. Lloyd, *Risks from ionising radiation: deterministic effects.* J Radiol Prot, 1998. **18**(3): p. 175-83.

21. Cox, R., H.G. Menzel, and J. Preston, *Internal dosimetry and tritium--the ICRP position.* J Radiol Prot, 2008. **28**(2): p. 131-5.

22. *The 2007 Recommendations of the International Commission on Radiological Protection. ICRP publication 103.* Ann ICRP, 2007. **37**(2-4): p. 1-332.

23. ICRP, *Relative Biological Effectiveness (RBE), quality factor (Q), and radiation weighting factor (w_R) ICRP Publication 92*. Ann ICRP. Vol. 33. 2004: Oxford: Pergamon.

24. Rooney, C., et al., *Case-control study of prostatic cancer in employees of the United Kingdom Atomic Energy Authority.* BMJ, 1993. **307**(6916): p. 1391-7.

25. Atkinson, W.D., D.V. Law, and K.J. Bromley, *A decline in mortality from prostate cancer in the UK Atomic Energy Authority workforce.* J Radiol Prot, 2007. **27**(4): p. 437-45.

26. Workers, I.S.G.o.C.M.a.N.I., *Direct estimates of cancer mortality due to low doses of ionizing radiation: an international study.* Lancet, 1994(334): p. 1039-1043.

27. (UNSCEAR), U.N.S.C.o.t.E.o.A.R., *Sources and effects of ionizing radiation.* 2000.

28. Richardson, D.B. and S. Wing, *Leukemia mortality among workers at the Savannah River Site.* Am J Epidemiol, 2007. **166**(9): p. 1015-22.

29. McGeoghegan, D. and K. Binks, *The mortality and cancer morbidity experience of employees at the Chapelcross plant of British Nuclear Fuels plc, 1955-95.* J Radiol Prot, 2001. **21**(3): p. 221-50.

30. Omar, R.Z., J.A. Barber, and P.G. Smith, *Cancer mortality and morbidity among plutonium workers at the Sellafield plant of British Nuclear Fuels.* Br J Cancer, 1999. **79**(7-8): p. 1288-301.

31. Zablotska, L.B., J.P. Ashmore, and G.R. Howe, *Analysis of mortality among Canadian nuclear power industry workers after chronic low-dose exposure to ionizing radiation.* Radiat Res, 2004. **161**(6): p. 633-41.

32. Cardis, E., et al., *The 15-Country Collaborative Study of Cancer Risk among Radiation Workers in the Nuclear Industry: estimates of radiation-related cancer risks.* Radiat Res, 2007. **167**(4): p. 396-416.

33. Dunson, D.B., *Commentary: practical advantages of Bayesian analysis of epidemiologic data.* Am J Epidemiol, 2001. **153**(12): p. 1222-6.

34. Greenland, S., *Bayesian perspectives for epidemiological research: I. Foundations and basic methods.* Int J Epidemiol, 2006. **35**(3): p. 765-75.

35. Greenland, S., *Bayesian perspectives for epidemiological research. II. Regression analysis.* Int J Epidemiol, 2007. **36**(1): p. 195-202.

36. MacLehose, R.F., et al., *Bayesian methods for highly correlated exposure data.* Epidemiology, 2007. **18**(2): p. 199-207.

37. Goodman, S.N., *Toward evidence-based medical statistics. 2: The Bayes factor.* Ann Intern Med, 1999. **130**(12): p. 1005-13.

38. Greenland, S., *Principles of multilevel modelling.* Int J Epidemiol, 2000. **29**(1): p. 158-67.

39. Gill, J., *Bayesian methods : a social and behavioral sciences approach*. 2nd ed. Statistics in the social and behavioral sciences series2008, Boca Raton: Chapman & Hall/CRC. xxxvii, 711 p.

40. Goodman, S.N., *Toward evidence-based medical statistics. 1: The P value fallacy.* Ann Intern Med, 1999. **130**(12): p. 995-1004.

41. Carlin, B.P. and T.A. Louis, *Bayesian methods for data analysis*. 3rd ed. Chapman & Hall/CRC texts in statistical science series2009, Boca Raton: CRC Press. xv, 535 p.

42. Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, *Markov chain Monte Carlo in practice*1998, Boca Raton, Fla.: Chapman & Hall. xvii, 486 p.

43. Heinze, G. and M. Schemper, *A solution to the problem of separation in logistic regression.* Stat Med, 2002. **21**(16): p. 2409-19.

44. Gelman, A., et al., *A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models.* Annals of Applied Statistics, 2008. **2**(4): p. 1360-1383.

45. Goodman, S.N., *Introduction to Bayesian methods I: measuring the strength of evidence.* Clin Trials, 2005. **2**(4): p. 282-90; discussion 301-4, 364-78.

46. Kadane, J.B., *Is "Objective Bayesian Analysis" objective, Bayesian, or wise? (Comment on Articles by Berger and by Goldstein).* Bayesian Analysis, 2006. **1**(3): p. 433-435.

47. Berger, J., *The Case for Objective Bayesian Analysis.* Bayesian Analysis, 2006. **1**(3): p. 385-402.

48. Chen, M.H., J.G. Ibrahim, and S. Kim, *Properties and Implementation of Jeffreys's Prior in Binomial Regression Models.* Journal of the American Statistical Association, 2008. **103**(484): p. 1659-1664.

49. Ibrahim, J.G. and P.W. Laud, *On Bayesian-Analysis of Generalized Linear-Models Using Jeffreys Prior.* Journal of the American Statistical Association, 1991. **86**(416): p. 981-986.

50. Spiegelhalter, D.J., et al., *Bayesian methods in health technology assessment: a review.* Health Technol Assess, 2000. **4**(38): p. 1-130.

51. Greenland, S., *Invited commentary: variable selection versus shrinkage in the control of multiple confounders.* Am J Epidemiol, 2008. **167**(5): p. 523-9; discussion 530-1.

52. Greenland, S., *Modeling and variable selection in epidemiologic analysis.* Am J Public Health, 1989. **79**(3): p. 340-9.

53. Weng, H.Y., et al., *Methods of covariate selection: directed acyclic graphs and the change-in-estimate procedure.* Am J Epidemiol, 2009. **169**(10): p. 1182-90.

54. Greenland, S., *Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators.* Biostatistics, 2000. **1**(1): p. 113-22.

55. Greenland, S. and J.M. Robins, *Identifiability, exchangeability and confounding revisited.* Epidemiol Perspect Innov, 2009. **6**: p. 4.

56. Greenland, S., J.A. Schwartzbaum, and W.D. Finkle, *Problems due to small samples and sparse data in conditional logistic regression analysis.* Am J Epidemiol, 2000. **151**(5): p. 531-9.

57. Heinze, G. and R. Puhr, *Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets.* Stat Med, 2010. **29**(7-8): p. 770-7.

58. Greenland, S., *Putting background information about relative risks into conjugate prior distributions.* Biometrics, 2001. **57**(3): p. 663-70.

59. Rappaport, S.M., et al., *The relationship between environmental monitoring and biological markers in exposure assessment.* Environ Health Perspect, 1995. **103 Suppl 3**: p. 49-53.

60. Renwick, A.G. and N.R. Lazarus, *Human variability and noncancer risk assessment--an analysis of the default uncertainty factor.* Regul Toxicol Pharmacol, 1998. **27**(1 Pt 1): p. 3-20.

61. Calabrese, E.J., B.D. Beck, and W.R. Chappell, *Does the animal-to-human uncertainty factor incorporate interspecies differences in surface area?* Regul Toxicol Pharmacol, 1992. **15**(2 Pt 1): p. 172-9.

62. Dunson, D.B. and B. Neelon, *Bayesian inference on order-constrained parameters in generalized linear models.* Biometrics, 2003. **59**(2): p. 286-95.

63. Vrijheid, M., et al., *The 15-Country Collaborative Study of Cancer Risk Among Radiation Workers in the Nuclear Industry: design, epidemiological methods and descriptive results.* Radiat Res, 2007. **167**(4): p. 361-79.

64. MacLehose, R.F., et al., *Bayesian methods for correcting misclassification: an example from birth defects epidemiology.* Epidemiology, 2009. **20**(1): p. 27-35.

65. Hamra, G., L.A. Nylander-French, and D. Richardson, *Dose reconstruction for an occupational cohort at the Savannah River nuclear facility: evaluation of a hybrid method.* Radiat Prot Dosimetry, 2008. **131**(2): p. 188-97.

66. Richardson, D.B., S. Wing, and S. Wolf, *Mortality among workers at the Savannah River Site.* Am J Ind Med, 2007. **50**(12): p. 881-91.

67. Richardson, D.B., S. Wing, and R.D. Daniels, *Evaluation of external radiation dosimetry records at the Savannah River Site, 1951-1989.* J Expo Sci Environ Epidemiol, 2007. **17**(1): p. 13-24.

68. Pierce, D.A. and D.L. Preston, *Radiation-related cancer risks at low doses among atomic bomb survivors.* Radiat Res, 2000. **154**(2): p. 178-86.

69. Straume, T. and A.L. Carsten, *Tritium radiobiology and relative biological effectiveness.* Health Phys, 1993. **65**(6): p. 657-72.

70. Gelfand, A.E., A.F.M. Smith, and T.M. Lee, *Bayesian-Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling.* Journal of the American Statistical Association, 1992. **87**(418): p. 523-532.

71. Greenland, S., *Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods.* Int J Epidemiol, 2009. **38**(6): p. 1662-73.

72. Maclehose, R.F. and D.B. Dunson, *Bayesian semiparametric multiple shrinkage.* Biometrics, 2010. **66**(2): p. 455-62.