Bayesian Approach to Estimating Reproductive Inhibition Potency In Aquatic Toxicity Testing

Jing Zhang[+*], A. John Bailer[+], and James T. Oris[++].

[+]Department of Statistics, Miami University, Oxford, Ohio, USA

[++]Department of Zoology, Miami University, Oxford, Ohio, USA

[*] To whom correspondence should be addressed (zhangj8@muohio.edu)

Abstract – Effectively and accurately assessing the toxicity of chemicals and their impact to the environment continues to be an important concern in ecotoxicology. Single experiments conducted by a particular laboratory often serve as the basis of toxicity risk assessment. These laboratories often have a long history of conducting experiments using particular protocols. In the present study, a Bayesian analysis for estimation of potency based on a single experiment was formulated, which then served as the basis for incorporating the experimental information from historical controls. A Bayesian hierarchical model was developed to estimate the reproductive inhibition concentrations (RIp) of a toxicant and flexible ways of using historical control information were suggested. The methods were illustrated using a data set produced by the test for reproduction in *Ceriodaphnia dubia* in which the number of young produced over three broods was recorded. In addition, simulation studies were included to compare the Bayesian methods with previously proposed estimators of potency. The Bayesian methods gave more precise RIp estimates with smaller variation and nominal coverage probability offsetting a small negative bias in the point estimate. Incorporating historical control information in the Bayesian hierarchical model effectively utilizes the useful information from past similar experiments when estimating the RIp, and results in potency estimates that are more precise compared to frequentist methods.

Keywords –Bayesian Hierarchical Models, Reproductive Toxicology, Historical Control Information; *Ceriodaphnia dubia*.

INTRODUCTION

Aquatic toxicology experiments are conducted to evaluate the potential impact of chemicals in receiving waters, marine systems, and other ecosystems. A single experiment typically includes a gradient of concentrations. Based on this experiment, an effective concentration associated with a specified degree of impact to an ecologically relevant response can be derived. A laboratory that conducts these experiments often conducts a large number of studies in a particular year. The objective of the present study is to construct a Bayesian potency estimator and describe how to incorporate the historical control information in such an estimate.

Reproduction responses are important endpoints in aquatic toxicity assessments and relative inhibition concentration (RIp), the concentration associated with a specific level of inhibition relative to control results, are often estimated. Potency estimation has been discussed in many situations. For example, linear regression or generalized linear regression models are fit to concentration-response data and these fitted models are inverted to produce estimates of the potency of toxic chemicals, e.g., the concentration of a chemical that produces some specified level of adverse response. For example, based on a single laboratory experiment, a generalized linear regression model was proposed to estimate the relative inhibition concentration [2,3]. Their proposed analysis was based on frequentist methods which provided the point estimates and confidence intervals for the RIp.

Bayesian methods have been previously applied in aquatic toxicology. For example, the no effect concentration was estimated via a Bayesian modelling approach [5]. Besides the flexibility shown in the model output provided by the Bayesian methods, it is also possible to incorporate different levels of variability into a Bayesian model utilizing a hierarchical framework and hence improve the modeling and potency estimation by using data arising from multiple studies. Laboratories that conduct many experiments over the course of a year or the experiments conducted by different laboratories over a long time period can be analyzed using a Bayesian hierarchical model to address the lab/source variability. Examples of such analyses from environmental toxicology and risk assessment include a model to analyze continuous data from multiple sources [4]

and a model to analyze non-normal data from multiple sources and provided the benchmark dose estimates [11].

Although Bayesian methods provide more information for decision making than frequentist methods (the posterior distributions rather than point estimates with error and confidence intervals that reflect probability statement about parameters) and could be adapted to model complicated data in different fields, there are some practical issues of doing Bayesian analysis. One of the concerns about the use of Bayesian analysis is the specification of prior distributions. Although one of the advantages of using Bayesian methods is the ability to incorporate any existing or "expert" knowledge in the practioner's field, we are aware of little discussion in aquatic toxicology research about incorporating such information into analysis via either modelling or prior specification. Our goal is to provide an example illustrating how to incorporate such historical information into analysis such that more practitioners could use Bayesian methods in their own toxicity assessment problems.

The *Ceriodaphnia dubia* reproduction toxicity test has been widely used in assessing the toxicity of effluents, environmental samples and single chemicals [12]. In the present study, we analyze the relationship between number of young produced and toxicant concentration, and then estimate the concentration associated with a specified decrement in the number of young produced in three broods. First we introduce the Bayesian framework of potency estimation. We then discuss how to utilize different types of historical information to improve our potency estimation. In the data analysis, we compare the results of frequentist method and Bayesian methods with historical information incorporated in different ways. In addition to applying these methods to real data, a small simulation study is included to illustrate the benefit of these methods.

METHODS

A Bayesian analysis begins with a formal statement about uncertainty in model parameters [7]. While a frequentist framework typically views parameters as unknown population constants, the Bayesian framework views parameters as random variables whose distributions can be better characterized given the observed data. Here, prior belief of the distribution of a parameter is

specified first, the "prior distribution." Data are used to update the prior belief about uncertainty and a distribution of the parameters given the data, the "posterior distribution", can be obtained. Historical information can be incorporated in the prior specification. In the present study, we describe the frequentist and Bayesian models for potency estimation and the incorporation of historical information in the estimation process. In this analysis, the relative inhibition estimator (RIp) is recast in a Bayesian formulation.

Data and notation

As a motivating example, forty-six *Ceriodaphnia dubia* reproduction toxicity tests were carried out in one lab during different time periods between July 18, 1989 and May 23, 1992. This data set is part of the data described and analyzed in Bailer et al. [1] and Wheeler et al. [11]. The experiments tested the impact of seven different levels (including control) of sodium chloride (NaCl) exposures on the reproduction of *Ceriodaphnia dubia*. The number of total young from three broods was recorded from organisms assigned to each concentration treatment as the response of interest. Histogram of the number of offspring of each organism in the control group is shown in **Figure 1**. Details of data layout are provided in **Table 1**, and notation for the components of this analysis is given below:

Y_{ij} : number of total young produced in three broods by the jth organism exposed to concentration c_i, where j = 1, 2, … , n_i and i =0, 1, …,g. Similarly, Y_{ij}^h is used in **Table 1** to denote the jth observation of number of young in the concentration c_i group recorded in the h th historical experiment (the superscript "h" denotes the h th historical experiment). Assuming there were H historical experiments in all, then $h = 1, 2, ..., H$.

μ_i : mean total young produced in three broods of organisms exposed to concentration c_i ; μ_0 is the mean total young produced in three broods with zero toxicant concentration, i.e., the control group mean.

116 β_k , k = 0, 1, 2, …, m: coefficients associated with the (function of) toxicant concentration
117 levels (m < number of concentration levels tested).

118 ***Estimation of Potency from a frequentist perspective***

119 The Poisson distribution is commonly assumed when modelling count data and was used to
120 derive a RIp estimate using the following generalized linear model proposed in [3]:

121 $$Y_{ij} \mid \mu_i \overset{independent}{\sim} Poisson(\mu_i) \quad (1)$$

122 $$\log(\mu_i) = \beta_0 + \beta_1 c_i + \beta_2 c_i^2 + ... + \beta_m c_i^m \quad (2)$$

123 Here the vertical bar "|" describes the distribution of the quantity to the left of the "|" given
124 information to the right; β_0 represents the intercept, and $\mu_0 = \exp(\beta_0)$ corresponds to the mean total
125 young in the control group; the other parameters, $\beta_1, ..., \beta_m$ are regression coefficients describing the
126 relationship between the mean total young and a function of toxicity concentration levels. Often
127 $m \leq 2$ is sufficiently flexible to model the toxicology data, i.e.,

128 $$\log(\mu_i) = \beta_0 + \beta_1 c_i + \beta_2 c_i^2 \quad (3)$$

129 Point estimates for the regression coefficients along with confidence intervals can be obtained using
130 likelihood-based inference methods. Given these parameters, the RIp is defined as the value of a
131 concentration which satisfies

132 $$\mu_{RIp} = (1-p)\mu_0 \quad (4)$$

133 where p is the proportion of inhibition and 0 < p < 1. For example, RI_{25} is the concentration where
134 $\mu_{RI_{25}} = 0.75\mu_0$ i.e. the concentration associated with a 25% decrement relative to the control mean.
135 Here, the RIp estimates are obtained as a function of the regression coefficients and p, $\widehat{RIp} =$
136 $f\left(p, \hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m\right)$. An approximate confidence interval of the RIp quantities can be derived from
137 the delta method [17] or bootstrapping [16].

Baysian reformulation of RIp estimation

Bayesian analysis is a statistical analysis strategy which treats a vector of parameters θ as unknown random variables instead of unknown constants, and therefore estimates the parameters of an underlying distribution based on the observed data **Y**. In our problem, $\theta = (\beta_0, \beta_1, ..., \beta_m)$ and $Y = (Y_{01}, ..., Y_{0n_0}, ..., Y_{g1}, ..., Y_{gn_g})$. The analysis begins with proposing a "prior distribution" for the parameters, denoted as $f(\theta)$. In the absence of strong prior beliefs, it is common to assume a uniform distribution over the appropriate range of values or other flat distributions for the prior distribution [7]. Flat prior distributions are distributions that have large variance and hence have density curves which are close to uniform density curve (flat straight line) over the appropriate range of parameter values.

Combining the prior distribution $f(\theta)$ and likelihood of the observed data $f(Y|\theta)$ using Bayes theorem, we update our knowledge about the parameter distribution, multiplying the likelihood function of sampling distribution by the prior distribution, and normalize to obtain a unit probability over all possible values

$$f(\theta|Y) = \frac{f(\theta) f(Y|\theta)}{f(Y)}$$

where $f(\theta|Y)$ is the posterior distribution of the parameters given the data and $f(Y)$ is the normalizing constant (see [7] for details). The mode or mean of the posterior distribution is often used as the Bayesian point estimate of a parameter while interval estimates are based on the quantiles of this distribution [7].

In the frequentist formulation of the RIp estimation, the Poisson likelihood and generalized linear model in **Equations 1-3** is the model for the total young count (Y_{ij}) given the regression parameters $\beta_0, ..., \beta_m$. In the Bayesian formulation, we assign prior distributions to these model parameters and generate samples of the regression parameters from their posterior distributions. Following the analyses done in Bailer et al. [3] and Wang et al. [10], we restrict ourselves to the case

where $m \le 2$. We start the Bayesian formulation assuming an independent normal distribution for the prior distribution of each of the regression coefficients

$$\beta_i \sim N(\beta_i^0, \sigma_i^2) \quad (5)$$

where i=0, 1, 2. Here β_i^0 is the prior mean of the coefficient β_i and σ_i^2 is the prior variance. The prior variance reflects our uncertainty in the parameter. When no prior information is available, a flat prior would be used for β_i, which means, the prior variance σ_i^2 is given as a fixed large number, or in a hierarchical setting, the uncertainty in the parameters is accounted for with the following higher level priors

$$\sigma_i^2 \sim Inv-Gamma(0.001, 0.001) \quad (6)$$

where $Inv-Gamma(0.001, 0.001)$ implies that the variance of the prior distribution of β_i, σ_i^2, is also unknown but has a particular assumed distribution (inverse gamma distribution). Inverse gamma distribution, denoted as $Inv-Gamma(a,b)$, is a widely used family of two-parameter distribution of positive continuous random variables and is the distribution of the reciprocal of a variable following the gamma distribution with the same parameters, $Gamma(a,b)$. Note that σ_i^2 cannot be negative or zero since it represents the uncertainty of our prior belief of the distribution of regression coefficient β_i ; therefore the inverse gamma distribution is usually used to assign prior distribution of variance parameters since it describes the distribution of positive continuous random variables. As discussed in [19], when we let a and b to be equal and small, the resulting inverse-gamma distribution then specifies a noninformative (flat) prior for variance parameters. Here we let $a = b = 0.001$, then the resulting flat prior of the variance parameter is $Inv-Gamma(0.001, 0.001)$. The prior distributions of regression coefficients allow for additional variability in the response, and therefore allow for over-dispersion, i.e., when the variance of count response is greater than the mean. **Table 2** presents a summary of this information for β_0 prior specification.

Since the regression coefficients $(\beta_0^0, \beta_1^0, \beta_2^0)$ are parameters in a higher level of the Bayesian hierarchical model shown in Equations 1, 2, 3 and 5, they are called “hyperparameters” and

need to be specified. Note that in the absence of knowledge about the concentration-response relationship, we assume flat priors for model parameters: in **Equation 5**, we use $\beta_i^0 = 0$; and flat priors with large prior variance, such as the inverse-gamma distributions given in **Equation 6** are used for the variance parameters σ_i^2, where i = 0, 1, 2. In a Bayesian analysis with non-informative priors, the data will strongly influence the analysis and results similar to a frequentist analysis are commonly observed. Note that it is possible to modify the flat priors and consider a different prior mean (β_0^0) for the intercept β_0. Since response levels in control conditions often need to be at least a certain amount for an experiment to pass quality assurance requirements and be used for potency estimation, it is reasonable to selected β_0^0 that satisfies that the expected number of offspring of organism in the control group specified by this prior, $\exp(\beta_0^0)$, is greater than or equal to the required control mean response.

Once the model parameters $(\beta_0, \beta_1, \beta_2)$ are sampled from their posterior distributions, we can calculate RIp based on these posterior samples. We compute a RIp for each sampled $(\beta_0, \beta_1, \beta_2)$ from the posterior distribution. Therefore, the implementation of our Bayesian model generates a distribution of the RIp that can provide a point estimate (e.g., posterior mean/mode) or an interval estimate and any quantiles of interest. Note that the Bayesian interval estimate is called a "credible interval", which is an analog of frequentist confidence intervals (CI), but unlike the frequentist CI, it is a probability statement which relates the intervals with probabilities. The frequentist confidence interval refers to the frequency that the confidence interval contains the parameter if the experiment is repeated. The Bayesian credible interval is an interval in the domain of a posterior probability distribution, and therefore for a given credible interval, one can directly compute the probability that the model parameter is in this interval by computing the area under the posterior density between the lower boundary and upper boundary of the credible interval (see more details in [7] and a comprehensive comparison in [15]).

Incoporating historical control information in to the estimation of potency

When little information is available for the regression coefficients and variances, noninformative (flat) priors would be used in the analysis. However, in a particular lab, there will often be additional information on the responses observed in control conditions. Since the comparison of reproduction ability depends on the evaluation of the control group, it is a natural choice for us to generate prior information for the control group based on the historical data, that is, utilizing the previous experimental results to specify the distribution of β_0. In this section, we illustrate how to derive informative priors for β_0 when the historical control information is available according to different scenarios.

Case 1: Summary statistics for historical controls are available. In this case, we examine how a prior for β_0 might be specified given only summary historical control information. The summary statistics available from the historical control data usually include the sample average and sample variance, the average of total young and the variation among the number of offspring produced by organisms in the control groups of previous experiments. As discussed before, $\mu_0 = \exp(\beta_0)$ corresponds to the mean total young in the control group, and therefore based on summary historical control data, we could generate a reasonable informative prior distribution for β_0. Suppose for the number of young observed in H historical control experiments with sample sizes of $n_0^1, n_0^2, ..., n_0^H$, the sample averages were recorded as $\overline{Y_0^1}, \overline{Y_0^2}, ..., \overline{Y_0^H}$, and sample variances were recorded as $S_{Y_0^1}^2, S_{Y_0^2}^2, ..., S_{Y_0^H}^2$. Based on these sample summary statistics, we could compute the overall sample average, $\overline{Y_{h0}} = \frac{n_0^1\overline{Y_0^1} + n_0^2\overline{Y_0^2} + ... + n_0^H\overline{Y_0^H}}{n_0^1 + n_0^2 + ... + n_0^H}$, and the pooled sample variance [18], $S_{Y_{h0}}^2 = \frac{(n_0^1 - 1)S_{Y_0^1}^2 + (n_0^2 - 1)S_{Y_0^2}^2 + ... + (n_0^H - 1)S_{Y_0^H}^2}{n_0^1 + n_0^2 + ... + n_0^H - H}$. These provide estimates for the mean and variance of the prior distribution for $\mu_0 = \exp(\beta_0)$ respectively. Here, we estimate the prior mean of μ_0, $E(\mu_0)$, with the overall sample average, $\overline{Y_{h0}}$, and the prior variance of μ_0, $V(\mu_0)$, with the pooled

sample variance, $S^2_{Y_{h0}}$. We could transform back and obtain the prior mean and variance of β_0. Since $\beta_0 = \log(\mu_0)$, the prior mean of β_0, $\beta_0^0 = \log[E(\mu_0)]$ is set equal to $\log(\overline{Y_{h0}})$; and we obtain the prior variance using the delta method [17], $\sigma_0^2 = \frac{V(\mu_0)}{E^2(\mu_0)}$ is set equal to $\frac{S^2_{Y_{h0}}}{(\overline{Y_{h0}})^2}$. Therefore, once we compute the estimated values for β_0^0 and σ_0^2, we could remove the higher level prior on σ_0^2 (Eqn. 6), and replace the flat prior distribution of β_0 specified in the hierarchical model (Eqn. 5) with the one derived from summary historical control information

$$\beta_0 \sim N(\log(\overline{Y_{h0}}), \frac{S^2_{Y_{h0}}}{(\overline{Y_{h0}})^2}) \quad (7)$$

Case 2: Actual historical control responses of individual organisms are available. When detailed experimental results are available for the historical control data, it is possible to implement a reduced Bayesian hierarchical model to analyze the historical control data only. When we look at the observations from the historical control group, $Y^1_{01},...,Y^1_{0n_0^1}, Y^2_{01},...,Y^2_{0n_0^2},....Y^H_{01},...,Y^H_{0n_0^H}$, the likelihood described in Equations 1and 2 reduces to

$$Y^k_{oj} \mid \mu_0 \overset{independent}{\sim} Poisson(\mu_0),\ j = 1,2,...,n_0^k; k = 1,2,...,H \quad (8)$$

$$\log(\mu_0) = \beta_0 \quad (9)$$

Then we simulate the posterior samples of β_0 from the posterior distribution of this reduced model using the flat prior of β_0 given in the previous section, e.g., $\beta_0 \sim N(0, \sigma_0^2)$ and $\sigma_0^2 \sim Inv-Gamma(0.001, 0.001)$. The resulting posterior samples serve as useful prior information for β_0 in the complete model when we analyze the current experiment results, providing information about the center, variation and shape of the prior distribution for β_0. From the histogram and density plot of the simulated posterior samples of β_0, we could decide which distributional family should be used to specify the prior distribution of β_0 based on the historical control information. If the histogram and density plot show that the historical control data suggests a symmetric bell-shaped

distribution for β_0, then it is reasonable to use a normal prior for β_0 in the complete model, and the hyper-parameters (β_0^0, σ_0^2) are decided using the summary statistics (sample mean and sample variance) of the posterior samples of β_0 from this reduced model.

Case 3: Actual historical concentration-response data using similar chemicals are available. Although different concentration levels are used in different experiments, when similar chemicals are used, the control groups among different experiments are expected to produce similar reproductive responses. In Case 2, we introduced how to produce prior distribution of β_0 based on the actual historical control data. An alternative way of utilizing the historical information is to implement the Bayesian hierarchical modelling on a combined data set which combines the information from the "current" data set and historical control data. In other words, the historical control data are not used as a prior-generation data set here, and is actually part of the model-building data set. When the historical control data are combined with the data produced by the current experiment, we have a larger sample size for the data, and will have more information to estimate the relationship between reproduction and concentrations, especially for the control group. Here we use flat prior for β_0, but we would expect to see the estimation of model parameters (especially β_0) affected by the incorporation of the historical control data. This case is analogous to a meta-analysis of the concentration-response analysis of a particular chemical.

To summarize the three possible methods of using the historical information, in Case 1 and 2, we generated the prior mean (β_0^0) and prior variance (σ_0^2) based on the availability of the historical control data; while in Case 3, we suggested combining the historical control data with the current experiment results for the analysis and use the flat priors. The useful historical information is incorporated into the Bayesian analysis either from the prior information or from the data. These analytic options are summarized in **Table 6** and will be described in more detail in Section 3.2.

Computational methods

WinBUGS [8] is a software package that makes practical Markov Chain Monte Carlo (MCMC) methods available to Bayesian methods users. Markov Chain Monte Carlo (MCMC) methods are used to obtain the samples from posterior distributions of the model parameters. R is a programming language and software environment for statistical computing and graphics [13].The R package "R2WinBUGS" [9] enables R users to implement a Bayesian model in WinBUGS software and save the simulations in arrays for easy access in R. In this study, we utilized the "R2WinBUGS" package to conduct the MCMC simulations.

In the simulation study, 1,000 toxicity experiments were simulated and five different methods were applied to each data set: the frequentist maximum likelihood approach was implemented using the "glm" function in R. The MCMC simulation for each of the four Bayesian models were used to produced 8,000 samples from the posterior distributions of the model parameters after a burn-in period of 2,000 iterations. In practice, an initial portion (burn-in) of the posterior samples from the MCMC simulation is often discarded in order to ensure convergence of the posterior samples. Based on the history plots of simulated posterior samples (shown in web appendix), we used 2,000 burn-in cycles in the simulation. It took approximately 867 min on a 2.93 GHz Intel CoreTM i7 CPU with R and "R2WinBUGS" software. In the application study, for all the models utilizing the historical information in different ways that we specified in previous sections, it took less than 270 s for us to run 50,000 MCMC iterations with the first 10,000 as burn-in iterations. Here we keep every 5th simulated posterior sample, i.e., we used a thinning interval of 5, and discard others to reduce the autocorrelation among posterior samples since MCMC produces correlated samples. The posterior quantities of regression coefficients, variance parameters and RIp's were obtained from the posterior samples. See the appendix for the R and WinBUGs model code.

Simulation study

We conducted a simulation study to compare the Bayesian methods with the frequentist method. First, we simulated 1000 experiments, assuming that each experiment has 10 observations at

each of the five different toxicity concentration-level groups: 0, 0.25%, 0.5%, 1% and 2%. For the simulated data set, we selected the true values of regression coefficients as $\beta_0 = 3.4$, $\beta_1 = 0.16$ and $\beta_2 = -0.58$ The true RIp values associated with these parameters could be computed with **Equation 4**, and they are $RI_{25} = 0.86$ and $RI_{50} = 1.24$. As mentioned before in Bailer et al. [3], the values of β_0, β_1 and β_2 should satisfy that $\beta_0 = \log(\mu_0) > 0$ and $\beta_2 < 0$. So we selected the true values assuming that the mean number of young for organisms in the control group is 30, $\beta_0 = \log(30) = 3.4$, and picked the number of β_1 and β_2 satisfying this restriction and resulting RIp quantities which were similar to the RIp estimates produced in Wheeler et al. [11] (see Table II in [11] for details).

For each of the 1000 simulated data sets, we estimated RI_{25} and RI_{50}, using the frequentist method and the Bayesian model with different choices of priors. The frequentist Poisson regression model specified by **Equations 1** and **2** was implemented to each of the 1000 simulated data sets; the regression coefficients and RIp's were estimated using maximum likelihood estimation, and the confidence intervals (CI) of RIp's were obtained using the delta method.

To apply the Bayesian method, we assumed the same likelihood function specified by Equations 1 and 2, and the priors for β_1 and β_2 were specified with **Equations 5** and **6**, with prior means $\beta_1^0 = \beta_2^0 = 0$. Four different priors for the intercept β_0 were used in the simulation and they were summarized in **Table 2**. Three different flat priors were used, assuming that β_0 had a normal prior distribution with the variance parameter following an inverse-gamma distribution, $\sigma_0^2 \sim Inv-Gamma(0.001, 0.001)$, but had different prior means, ln(30), ln(20) or 0. The three different prior means specified in the flat prior for β_0 indicated different prior belief people had about the control group response. Here, ln(30) or ln(20)translated into a control group mean of 30 or 20; which corresponded to the case that people had partial information about the center of the distribution of β_0 but with large uncertainty. We also considered the commonly used non-informative prior mean for regression coefficients in such hierarchical modeling, $\beta_0^0 = 0$, which

translated to a control group mean of 1 and not likely to be a reasonable prior belief in the experiments described here. The three different prior means were used in the simulation studies to check the sensitivity of choice of prior means in the Bayesian hierarchical model. Besides the flat priors, an informative prior for β_0, $\beta_0 \sim N(\ln(30), 0.001)$, with a very small variance (0.001) was also used, which represented the situation that people had strong prior belief about the control group mean response. This prior distribution indicated strong belief that the average reproduction count for control group should be very close to $\mu_0 = e^{3.4} = 30.$ For each of the four different prior specification for β_0, we simulated the regression coefficients and RIp's from their posterior distributions and used posterior means as estimates of the RIp in the population. The CIs of RIp's were computed based on the posterior sample quantiles.

In this simulation study, we compared the frequentist method (no prior input), Bayesian method with flat priors for all non-intercept regression coefficients and for the intercept: either weak prior input with the prior mean of β_0 far from the true value: $\beta_0^0 = 0 \neq \ln(30)$, weak prior input with the prior mean of β_0 close to the true value: $\beta_0^0 = \ln(20)$, weak prior input with the prior mean of β_0 equal to the true value: $\beta_0^0 = \ln(30)$, or informative prior for β_0 (strong prior input for the control group). We are able to evaluate the benefits of incorporating correct and strong prior information into analysis through the comparison of model performance in the simulation.

In the simulation study, the point estimates and standard deviations of the RIp's were computed for each simulated data set, using the frequentist method and each one of the four Bayesian methods with different priors for β_0. Then the average of points estimates, the square-root of the mean-squared error in RIp estimates (RMSE) and the observed percentage of nominal 95% coverage probability were computed for each method. Note that the coverage probability is the percentage of the 1000 resulting interval estimates of RIp's that contain the true RIp values. A good statistical model is expected to produce unbiased point estimate (the average of points estimates close to true values), smaller overall variability and bias in estimation (smaller RMSE) and coverage

probabilities close to the nominal specified levels (95% coverage probability close to 95%). The histograms of the 1000 RIp estimates produced by each method were plotted to compare the performance of different methods. Various percentiles of the 1000 estimates of RIp were also computed for each method applied in the simulation study.

We repeated the simulation study for other experimental scenarios, with five or three animals instead of ten, in order to examine the impact of sample size on estimation of RIp. Similarly, the point estimates and standard deviations of the RIp's were computed for each simulated data set, using the frequentist method and Bayesian methods with different priors for β_0: And the average of points estimates, the **RMSE** in RIp estimates, the percentage of 95% coverage probability and the average length of 95% interval estimates (**AL**) were computed for each method.

RESULTS

Simulation results

The simulation results were summarized in **Tables 3-5**. **Table 3** and **Table 4** presented those in which 10 animals were simulated for each concentration group. Although the simulation results in **Table 3** and **Table 4** showed that the average of the 1000 point estimates given by the frequentist method was closest to the true RIp values, but frequentist method showed the largest overall estimation error measured by **RMSE,** the frequentist percentiles of RIp estimates showed the widest range, and the coverage of the frequentist CI based on the delta method did not achieve nominal coverage levels of 95%. All the Bayesian methods performed similarly in terms of the bias in estimating RI_{25} and RI_{50} with average of points estimates slightly smaller than the true RIp values, but they all provided interval estimates that achieved the nominal coverage probabaility of 95%. In **Figure 2**, we display histograms of the point estimates of RI_{25} and RI_{50} based on the five methods. To make the comparison easier, we plotted the kernel density estimates [14] of RIp's for all five methods in the same graph, as shown in **Figure 3** and **Figure 4**. The histograms and density curves described the distributions of the 1000 RIp estimates produced by different methods. They showed

379 that Bayesian methods, especially the one using informative priors, produced a narrower range in
380 point estimates compared to the frequentist method. The Bayesian method with an informative prior
381 provided more precise RIp estimates (smaller **RMSE**) compared to the flat prior model because it
382 incorporated useful prior information and enhanced the updated knowledge of RIp quantities. The
383 results also showed that the Bayesian methods using a flat prior for β_0 but centered at different
384 location (ln(20) or ln(30)) gave similar results (the average of points estimates, **RMSE** and **the**
385 **coverage probability**) to the model with a flat prior for β_0 centered at 0; however, as shown in
386 **Figures 2, 3** and **4**, the center of the resulting populations of RIp estimates of these Bayesian
387 methods were moving towards the true values as the prior mean of β_0 increased from 0 to ln(30).
388 This happened because the incorporation of the partial prior information about responses in the
389 control group. Also, the Bayesian methods tended to underestimate the RI_{25} and RI_{50}, with the
390 distributions of RI_{25} and RI_{50} not centered at the true value. Based on this small simulation study, the
391 Bayesian method with an informative prior for β_0 was able to produce more precise point estimates,
392 albeit with small bias, that provided interval estimates which achieved nominal coverage
393 probabilities.

394 The results of small sample size simulation studies are summarized in **Table 5**. Frequentist
395 method gave unbiased values of the average of points estimates, while all the Bayesian methods gave
396 negatively biased values of the average of points estimates (i.e. underestimate true RIp on average)
397 but close to the true values. All the Bayesian interval estimates achieved the nominal CP of 95%
398 while the frequentist method failed to do so for RI_{25} estimation with smaller sample sizes. Further,
399 the Bayesian methods, especially the one with informative prior for β_0, gave much smaller average
400 length of interval estimates and much smaller **RMSE** for both RI_{25} and RI_{50}. Based on the results
401 shown in **Tables 3-5**, we could see that although the Bayesian methods slightly underestimated the
402 RIp point estimate, the resulting error in the point estimate (**RMSE**) was consistently smaller than
403 the frequentist method. Simultaneously, the Bayesian interval estimates were narrower (shorter
404 average length of interval estimates) and achieved the nominal coverage probability, and hence more

informative and more reliable. Among the Bayesian methods implemented in the simulation study, the Bayesian method with informative priors gave smallest estimation error (**RMSE**) and shortest average length of interval estimates for RIp. The Bayesian methods using a flat prior for β_0 but using ln(20) or ln(30) as prior means gave slightly smaller average lengths of interval estimates and **RMSE**s compared to the Bayesian method using a zero prior mean for β_0, but were not as good as the analysis using informative priors. As expected, in the small sample scenarios, the Bayesian methods using informative priors gave reasonable point estimates with smaller errors and narrower interval estimates achieving nominal coverage probabilities compared to the other methods. It mitigated the effects of small number of replicates and fully utilized the available historical information. Therefore, incorporating "correct" prior information, e.g., historical information, into the analysis led to more precise potency estimates.

Application

A collection of experiments conducted by a number of labs was described in Bailer et al. [1], including the reproduction and survival responses of *Ceriodaphnia dubia*. In this section, we used the experimental results of the reproduction response for *Ceriodaphnia dubia* tested in a lab on May 23, 1992 as the "current" data set to illustrate how to incorporate historical information into a Bayesian hierarchical model for toxicity assessment. In the "current" experiment, the observed number of total young produced in three broods was recorded for 70 organisms assigned to g = 7 different toxicity-level exposition groups. The seven different concentration levels of toxicant used in the experiment were: 0 (control group), 0.063%, 0.125%, 0.25%, 0.5%, 1% and 2%. For each concentration level, there were 10 observations. In the 45 historical experiments conducted before May 23, 1992, initially 10 organisms were used in each concentration group, which would produce 450 historical control observations; however, there were missing values in the number of young recorded for two of these experiments, and 7 instead of 10 organisms were observed in the control group. Therefore, in the H=45 historical experiments conducted by this lab, there were

(n1+n2+…+nH)=444 observations available in the control group and additional 2330 observations available in different toxicity level groups. **Figure 1** displays histogram of the total number of young per control group in all the historical experiments and the current experiment, which shows the observed distribution of reproduction in the control group.

Table 6 summarized the different methods we compared in the application study. First, we began with the frequentist method described in Bailer et al. [3], computing the maximum likelihood estimates for the regression parameters. The RIp point estimates were obtained using the method developed in Bailer et al. [3] and the 95% confidence intervals for RIp were obtained via the delta method[17]. We also implemented the Bayesian model with flat priors (Bayesian model 1 in **Table 6**) to compare with the frequentist method, expecting to obtain similar reproduction inhibition estimates since no information was given in the priors. Then we implemented the Bayesian methods with informative priors based on the historical control information. To illustrate the use of the historical information proposed in Section 2, we implemented the Bayesian hierarchical modelling framework described in Section 2.2 and 2.3 under different scenarios respectively, i. e. incorporating the summary or actual historical control information in the prior generation or incorporating the actual historical control information into the model-building data set.

We obtained an informative prior for regression coefficient "β_0" based on the summary historical control information (Bayesian model 2 in **Table 6**). Here summary statistics of the historical control group observations are used to specify the prior distribution for β_0. The 444 control group observations produced in 45 historical experiments had an overall mean of 29.3 and a pooled variance of 70.64; which translated into a prior distribution of β_0 as $\beta_0 \sim N(3.38, 0.08)$ when the mean and variance are incorporated into **Equation 7**.

We also implemented the model described in case 2 of Section 2.4, generating a sample of β_0 from the posterior distribution based on the historical control data first, and then studied its distribution characteristics. **Figure 5** showed that the posterior distribution of β_0 given the historical control data was approximately symmetric. The histogram and the density curve both confirmed that

it was reasonable to assume a normal prior for the regression coefficient β_0. The sample average and variance of the posterior sample served as a natural choice for the prior mean and variance when we incorporated the historical control information into the hierarchical model to analyze the current data (Bayesian model 3 in **Table 6**). The resulting prior of β_0 for this method was: $\beta_0 \sim N(3.38, 0.0001)$. Lastly, we combined the 444 historical control group observations into the model-building data set and implemented the hierarchical model again with at priors for all the parameters (Bayesian model 4 in **Table 6**), although not much differences were shown in the results with Bayesian model 4 relative to Bayesian model 3.

Table 7 summarizes the point estimates (**PE**), standard deviations (**SD**) and 95% confidence interval (for the frequentist method) and 95% credible intervals (for the Bayesian methods) of the RIp's (more specifically, p = 0.25 and p = 0.50) and regression coefficients. Note that the same group of reproduction responses recorded in the "current" experiment conducted on May 23, 1992 was used as the current experimental data set in the frequentist method and the first three Bayesian methods (using flat prior, prior generated from the summary historical control information, and prior generated from actual historical control observations for β_0). The estimates of regression coefficients β_0, β_1 and β_2 showed some differences in **Table 7**: the frequentist method and the Bayesian method 1 (using flat priors) or Bayesian model 2 (using informative prior for β_0 based on the summary statistics) gave very similar PE, SD and 95% CI; however, as shown by the 95% interval estimates, the Bayesian model 3 (using informative prior for β_0 generated from actual historical control observations) gave significantly larger estimated values of β_0 and β_2, together with a significantly smaller estimated value of β_1. In addition, the standard errors of the parameters ($\beta_0, \beta_1, \beta_2$) are smaller when using informative priors vs. flat priors. The 95% CI for β_1 did not cover 0 for the frequentist method ([0.17, 0.78]), the Bayesian model 1 ([0.04, 0.71]) or the Bayesian method 2 ([0.04, 0.73]), which showed that these models agreed that β_1 was significantly different from zero and it was necessary to include the first order term "$\beta_1 c_i$" in Equation 2 based on the

model output. However, the 95% CI ([-0.29, 0.08]) given by the Bayesian model 3 suggested that β_1 was not significantly different from zero. It was no surprise that the Bayesian model 1 (using flat priors) and frequentist method gave similar answers since they were utilizing the information from only the current data. In addition, using the prior generated from the summary historical control information (Bayesian model 2) had less impact on the output compared to the prior generated using all the actual historical control observations (Bayesian model 3), because the first method generated the prior based on the sample mean and variance of the historical control data only, while the second method generated the prior using the actual historical data, more information which translated into much stronger prior belief about β_0 with a much smaller prior variance used (0.0001 compared to 0.18). This suggests that when the prior belief about the parameters incorporated into the analysis are not strong (having comparably large prior variances), the Bayesian model would produce similar results as those resulted by flat priors.

The Bayesian model 3 showed obvious differences in either RIp or parameter estimation compared to the Bayesian model 1 or Bayesian model 2. Therefore, picking different priors for β_0 would affect the practioner's understanding of the reproductive- toxicity relationship and change the answers about RIp's. The estimated RIp's given by the Bayesian method 1or Bayesian model 2 were larger and closer to the frequentist results. The Bayesian model 3 allowed the prior information to be generated from the output of model described in **Equations 8** and **9**; therefore, it enhanced knowledge about β_0 and also improved the estimation of RIp quantities. As shown in the last column of **Table 7**, when the actual historical control observations were combined with the model-building data (“current” experiment data) in Bayesian model 4, similar model outputs were produced compared to the Bayesian model 3. This reflects that increasing the number of data observations in the control group when we combine the historical control data into the model-building data set directly in Bayesian model 4 generates similar information about the β_0 prior as incorporated into the model by generating the prior distribution of β_0 purely determined by the actual historical control data in Bayesian model 3.

The potency estimates, RI_{25} and RI_{50}, were strongly impacted by historical control information. The current experiment had a mean response in the controls that was much smaller than the historical controls. Incorporating the historical control information, informative priors brought in a strong prior belief that the control response should be higher than the current response, and enables the inhibition to be estimated relative to this higher value. This translated to smaller RI_{25} and RI_{50} estimates when control information was incorporated. So discussion of how much weight should be placed on historical data should be made before analysis.

When incorporating the historical control information in the analysis of a current data set, the past experiments should be examined to confirm that they were conducted under similar conditions and could be reasonably considered as a sample from the same population. When the historical control data are "comparable", the resulting parameter estimates benefit from incorporating the historical control information; however, if the historical control data were not similar to the current experimental conditions, i.e., data were collected under different conditions, the resulting analysis would be misleading.

The model parameters ($\beta_0, \beta_1, \beta_2$) are estimated for each method and then used to estimate the expected number of young based on Equations 1 and 2 so that we could obtain the relationship between toxic concentration and expected number of young. Based on the different parameter estimates produced by different methods, the estimated concentration-response curves produced by the frequentist method and Bayesian methods were plotted in **Figure 6**. Note that only three different curves were shown in **Figure 6**, since the Bayesian models 1 and 2 generated almost the same estimated concentration-response relationship as determined by the similar estimated regression coefficients given in **Table 7**, and the Bayesian model 3 and 4 generated the same estimated concentration-response curve. All the three different curves showed quite similar trend except for the area with low toxicity exposure ($\leq$0.5%). Here the historical control information impacted the output of the zero/low toxicity directly in the Bayesian models, especially in Bayesian method with priors

generated on the actual historical observations and the one using combined data set, and therefore changed the estimation for this area and resulted the difference shown in the graph.

DISCUSSION

In the present study, we develop potency estimators based on a Bayesian perspective that can directly incorporate historical control information, using either the summary historical control information or the individual historical control experimental results. These methods summarize how to incorporate the useful information from previous studies and improve their understanding of current experiment results. The application example provides a template of doing Bayesian analysis and generating informative priors in an aquatic toxicology context.

Note that in the present study, we treat the historical control observations as if they were from the same population as the "current" control group data. It is also possible to use them with a "discount" when we know that they are from a group with similar population center but with different variability. We could inflate the prior variance for β_0 in that case by a constant to decrease the impact brought in by the prior information to be lowered. When one is not clear about the reliability of the historical information, a Bayesian model with flat priors is applicable and still outperforms the frequentist method in terms of coverage probabilities of interval estimates, although alternatives to the delta-method-based CI have been suggested for the frequentist RIp methods, e. g., parametric-bootstrap-based CI.

One attractive feature of using the Bayesian hierarchical modelling is that the Bayesian analysis results in probability statements about parameters. Thus, it is sensible to state that the probability that RI_{25} is between 0.87 and 1.24 is 0.95 from the perspective of a Bayesian analysis. Further, prior information such as historical information from a lab can be naturally incorporated. As shown in the simulation study, when one has reliable historical information which can be incorporated into the modelling, a Bayesian model with informative priors would be a better choice since it results smaller estimation error and narrower range of the interval estimates in RIp

estimation. Another important benefit of incorporating the historical information in the Bayesian hierarchical modelling is that fewer organisms may be needed in an experiment. It was shown in the simulation study with only three or five animals per concentration group that the interval estimates of RIp's given by the Bayesian method using informative priors gave much smaller range compared to the frequentist results and achieved the nominal coverage probabilities. This indicates the possibility of using reduced number of animals in toxicity assessment.

In future work, we might be able to modify our Bayesian hierarchical model to accommodate for the situation where the testing subjects are exposed to higher toxicity concentrations with increasing mortality. When the chemical concentration increases in experiments and exceeds a certain "threshold concentration", the probability of obtaining zeroes in the number of young produced by an organism increases dramatically. This happens because the high toxicity is usually associated with high mortality or low reproduction. In that case, the Poisson distribution might not be an appropriate distribution assumption anymore and a model with compatible parameter to evaluate the probability of observing zero young for a given toxicity centration is needed. For tests using high level toxicity exposure, we could possible modify our Bayesian hierarchical model and tailor it according to this change to improve our estimation of RIp. In addition to relating the number of young with the concentration, we would also consider relating the presence/absence of young with the concentration level. Zero-inflated Poisson models would be a possible choice to model such data set [10]. A Bayesian reformulation of this problem would be natural extension of our work.

REFERNCES

[1] Bailer, A.J., Hughes, M.R., Denton, D.L., Oris, J.T. 2000. An empirical comparison of effective concentration estimators for evaluating aquatic toxicity test responses. *Environ Toxicol Chem* **19**:141-150.

[2] Bailer, A.J., Oris, J.T. 1993. Modeling reproductive toxicity in Ceriodaphnia tests. *Environ Toxicol Chem* **12**: 787-791.

582 [3] Bailer, A.J., Oris, J.T. 1997. Estimating inhibition concentrations for different response scales
583 using generalized linear models. *Environ Toxicol Chem* **16**: 1554-1559.
584 [4] Coull, B.A., Mezzetti, M., Ryan, L.M. 2003. A Bayesian hierarchical model for risk assessment
585 of methylmercury. *Journal of Agricultural, Biological, and Environmental Statistics* **8**: 253-
586 270.
587 [5] Fox, D.R. 2010. A Bayesian approach for determining the no effect concentration and hazardous
588 concentration in ecotoxicology. *Ecotoxicology and Environmental Safety*, **73**: 123-131.
589 [6] Gelfand, A.E., Silander, Jr. J.A., Wu, S., Latimer, A., Lewis, P.O., Rebelo, A.G., Holder, M.
590 2006. Explaining species distribution patterns through hierarchical modeling. *Bayesian*
591 *Analysis* **1**: 41-92.
592 [7] Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. 2003. Bayesian Data Analysis, 2nd edition,
593 CRC Press, London.
594 [8] Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D. 2000. WinBUGS -- a Bayesian modelling
595 framework: concepts, structure, and extensibility. *Statistics and Computing* **10**: 325-337.
596 [9] Sturtz, S., Ligges, U., Gelman, A. 2005. R2WinBUGS: A Package for Running WinBUGS
597 from R. *Journal of Statistical Software* **12**: 1-16.
598 [10] Wang, S.C., Smith, E.P. 2000. Adjusting for morality effects in chronic toxicity testing:
599 Mixture model approach. *Environ Toxicol Chem* **19**: 204-209.
600 [11] Wheeler, M.W., Bailer, A.J. 2009. Benchmark dose estimation incorporating multiple data
601 sources. *Risk Analysis* **29**: 249-256.
602 [12] Weber, C.I., Peltier, W.H., Norberg-King, T.J., Horning, W.B., Kessler, F.A., Menkedick,
603 J.R., Neiheisel, T.W., Lewis, P.A., Kemm, K., Pickering, Q.H., Robinson, E.L., Lazorchack,
604 J.M., Wymer, L.J., Freyberg, R.W. 1989. Short-Term Methods for Estimating the Chronic
605 Toxicity of Euents and Receiving Waters to FreshwaterOrganisms. Cincinnati, OH: U.S.
606 Environmental Protection Agency, 2nd edition, EPA/600/4-89/001A
607 [13] R Development Core Team 2010. R: A Language and Environment for Statistical Computing.
608 Vienna, Austria, R Foundation for Statistical Computing.

609 [14] Silverman, B.W. 1986. Density Estimation. London: Chapman and Hall.

610 [15] Zech, G. 2002. Frequentist and Bayesian confidence intervals. *European Physical Journal*
611 *Direct.*

612 [16] Efron, B., R. Tibshirani. 1994. An Introduction to the Bootstrap, Chapman & Hall, London,
613 UK.

614 [17] Morgan, B.J.T. 1992. Analysis of Quantal Response Data, Chapman & Hall, London, UK.

615 [18] Ott, L., Longnecker, M., Ott, R. 2000. An introduction to statistical methods and data
616 analysis. New York: Duxbury.

617 [19] Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. 1994, 2003.
618 BUGS:Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge,
619 England. www.mrc-bsu.cam.ac.uk/bugs/

620

621

622

623 **Tables**

624 Table 1: Illustration of the lay out of the data

concentration	c_0	c_1	…	c_g
current experiment[a]	$Y_{01},...,Y_{0n_0}$	$Y_{11},...,Y_{1n_1}$	…	$Y_{g1},...,Y_{gn_g}$
historical experiment 1[b]	$Y^1_{01},...,Y^1_{0n^1_0}$	$Y^1_{11},...,Y^1_{1n^1_1}$	…	$Y^1_{g1},...,Y^1_{gn^1_g}$
historical experiment 2[c]	$Y^2_{01},...,Y^2_{0n^2_0}$	$Y^2_{11},...,Y^2_{1n^2_1}$	…	$Y^2_{g1},...,Y^2_{gn^2_g}$
…	…	…	…	…
historical experiment H[d]	$Y^H_{01},...,Y^H_{0n^H_0}$	$Y^H_{11},...,Y^H_{1n^H_1}$	…	$Y^H_{g1},...,Y^H_{gn^H_g}$

625

626 [a] $n_0, n_1, ..., n_g$ are the number of animals tested in the corresponding toxicity level groups of the current
627 experiment

[b] $n_0^1, n_1^1, ..., n_g^1$ are the number of animals tested in the corresponding toxicity level groups of the 1st historical experiment

[c] $n_0^2, n_1^2, ..., n_g^2$ are the number of animals tested in the corresponding toxicity level groups of the 2nd historical experiment

[d] $n_0^H, n_1^H, ..., n_g^H$ are the number of animals tested in the corresponding toxicity level groups of the H th historical experiment

Table 2: Prior distributions for β_0 used in the simulation study

Bayesian models	Prior distribution for β_0	Parameters used in the prior distribution	
		β_0^0	σ^2
Informative prior	$N(\beta_0^0,\sigma^2)$	3.4	0.001
Flat prior, centered at 0	$N(\beta_0^0,\sigma^2)$	0	σ^2~Inv-Gamma(0.001,0.001)
Flat prior, centered at ln(20)	$N(\beta_0^0,\sigma^2)$	ln(20)	σ^2~Inv-Gamma(0.001,0.001)
Flat prior, centered at ln(30)	$N(\beta_0^0,\sigma^2)$	ln(30)	σ^2~Inv-Gamma(0.001,0.001)

Table 3: Comparison of simulation study results (RI_{25}) when 10 organisms are simulated for each concentration group

Methods		RI_{25} (true value: 0.86)				
		Frequentist Method	Bayesian Methods With flat priors for β_1 and β_2 and different priors for			
			Flat prior, centered at 0	Flat prior, centered at ln(20)	Flat prior, centered at ln(30)	Informative prior
APE[a]		0.85	0.82	0.82	0.82	0.82
RMSE[b]		0.09	0.08	0.08	0.07	0.06
95%CP[c]		73.30%	94.50%	94.70%	95.70%	96.30%
RI_{25} percentiles[d]	1%	0.62	0.66	0.67	0.69	0.73
	2.5%	0.66	0.70	0.70	0.71	0.74
	5%	0.70	0.71	0.72	0.73	0.75
	10%	0.73	0.73	0.74	0.75	0.76
	25%	0.79	0.77	0.77	0.78	0.79
	50%	0.85	0.80	0.81	0.81	0.82
	75%	0.92	0.85	0.86	0.85	0.85
	90%	0.97	0.91	0.92	0.90	0.89
	95%	1.01	0.95	0.96	0.94	0.92
	97.5%	1.04	0.99	1.00	0.96	0.94
	99%	1.07	1.03	1.04	1.00	0.96

[a]APE=average of point estimates among 1000 simulations

[b]RMSE=the square root of the mean squared error

[c]95%CP=the percentage of 95% interval estimates covering the true values of parameters

[d]RI_{25} percentiles=percentiles of the 1000 RI_{25} estimates with the lower tail percentage specified as 1%, 2.5%, 5%, 10%, 25%, 50%, 75%, 90%, 97.5% and 99%.

Table 4: Comparison of simulation study results (RI_{50}) when 10 organisms are simulated for each concentration group

Methods		RI_{50} (true value: 1.24)				
		Frequentist Method	Bayesian Methods With flat priors for β_1 and β_2 and different priors for β_0			
			Flat prior, centered at 0	Flat prior, centered at ln(20)	Flat prior, centered at ln(30)	Informative prior
APE[a]		1.23	1.22	1.22	1.22	1.22
RMSE[b]		0.07	0.07	0.06	0.06	0.05
95%CP[c]		89.20%	95.00%	94.90%	96.20%	96.40%
RI_{50} percentiles[d]	1%	1.08	1.09	1.09	1.10	1.13
	2.5%	1.10	1.11	1.11	1.12	1.14
	5%	1.12	1.12	1.13	1.13	1.15
	10%	1.15	1.14	1.15	1.15	1.16
	25%	1.19	1.17	1.18	1.18	1.19
	50%	1.24	1.21	1.21	1.21	1.22
	75%	1.28	1.25	1.26	1.25	1.25
	90%	1.33	1.29	1.30	1.28	1.28
	95%	1.35	1.32	1.33	1.31	1.30
	97.5%	1.37	1.34	1.35	1.33	1.31
	99%	1.39	1.36	1.37	1.35	1.33

[a]APE=average of point estimates among 1000 simulations

[b]RMSE=the square root of the mean squared error

[c]95%CP=the percentage of 95% interval estimates covering the true values of parameters

[d]RI_{50} percentiles=percentiles of the 1000 RI_{50} estimates with the lower tail percentage specified as 1%, 2.5%, 5%, 10%, 25%, 50%, 75%, 90%, 97.5% and 99%.

Table 5: Comparison of simulation study results (RI_{25} and RI_{50}) when 3 or 5 organisms are simulated for each concentration group

Methods	RI_{25} (true value: 0.86)							
	APE[a]		RMSE[b]		95%CP[c]		AL[d]	
	n_i=3	n_i=5	n_i=3	n_i=5	n_i=3	n_i=5	n_i=3	n_i=5
Frequentist Method	0.85	0.86	0.17	0.13	86.9%	84.6%	0.64	0.41
Bayesian Methods								
Informative prior	0.81	0.82	0.09	0.07	97.0%	97.2%	0.42	0.31
Flat prior, centered at 0	0.79	0.81	0.13	0.10	97.9%	95.8%	0.65	0.43
Flat prior, centered at ln(20)	0.79	0.81	0.11	0.09	98.8%	97.2%	0.54	0.40
Flat prior, centered at ln(30)	0.81	0.82	0.13	0.10	98.0%	96.8%	0.63	0.44
Methods	RI_{50} (true value: 1.24)							
	APE*		RMSE**		95%CP+		AL++	
	n_i=3	n_i=5	n_i=3	n_i=5	n_i=3	n_i=5	n_i=3	n_i=5
Frequentist Method	1.24	1.23	0.13	0.1	98.2%	96.8%	0.80	0.50
Bayesian Methods								
Informative prior	1.24	1.26	0.08	0.06	97.3%	97.2%	0.34	0.27
Flat prior, centered at 0	1.21	1.21	0.11	0.08	95.8%	97.0%	0.50	0.36

Flat prior, centered at ln(20)	1.21	1.21	0.09	0.07	96.9%	98.0%	0.43	0.33
Flat prior, centered at ln(30)	1.22	1.22	0.11	0.08	96.1%	97.2%	0.48	0.36

[a]APE=average of point estimates of RIp's among 1000 simulations

[b]RMSE=the square root of the mean squared error in RIp estimates

[c]95%CP=the percentage of 95% interval estimates of RIp's covering the true values of parameters

[d]AL=the average length of 95% interval estimates of the RIp's.

Table 6: Illustration of different models used in the application study. The methods described in the table were applied to the real data produced by the *Ceriodaphnia dubia* tests conducted in lab "CAAQS" on May 23, 1992.

Model	Likelihood Function	Priors	Data Set Used
Frequentist model	$Y_{ij} \mid \mu_i \overset{independent}{\sim} Poisson(\mu_i)$ $\log(\mu_i) = \beta_0 + \beta_1 c_i + \beta_2 c_i^2$	Not available.	current experiment results
Bayesian Model 1 (flat priors, all centered at 0)	$Y_{ij} \mid \mu_i \overset{independent}{\sim} Poisson(\mu_i)$ $\log(\mu_i) = \beta_0 + \beta_1 c_i + \beta_2 c_i^2$	$\beta_i \sim N(\beta_i^0, \sigma_i^2)$, i=0,1,2 $\sigma_i^2 \sim Inv-Gamma(0.001, 0.001)$	current experiment results
Bayesian Model 2			current

(flat priors centered at 0 were used for β_1 and β_2, and informative prior was used for β_0, with prior mean and variance computed based on the summary historical control information)	$Y_{ij} \mid \mu_i \overset{independent}{\sim} Poisson(\mu_i)$ $\log(\mu_i) = \beta_0 + \beta_1 c_i + \beta_2 c_i^2$	$\beta_i \sim N(\beta_i^0, \sigma_i^2)$, i=1,2 $\sigma_i^2 \sim Inv-Gamma(0.001, 0.001)$ $\beta_0 \sim N(3.38, 0.18)$	experiment results
Bayesian Model 3 (flat priors centered at 0 were used for β_1 and β_2, and informative prior was used for β_0, with prior mean and variance computed based on the actual historical control information)	$Y_{ij} \mid \mu_i \overset{independent}{\sim} Poisson(\mu_i)$ $\log(\mu_i) = \beta_0 + \beta_1 c_i + \beta_2 c_i^2$	$\beta_i \sim N(\beta_i^0, \sigma_i^2)$, i=1,2 $\sigma_i^2 \sim Inv-Gamma(0.001, 0.001)$ $\beta_0 \sim N(3.38, 0.0001)$.	current experiment results
Bayesian Model 4 (flat priors, all centered at 0, but the data set used contains both current experiment results and previous experiment results in the control group)	$Y_{ij} \mid \mu_i \overset{independent}{\sim} Poisson(\mu_i)$ $\log(\mu_i) = \beta_0 + \beta_1 c_i + \beta_2 c_i^2$	$\beta_i \sim N(\beta_i^0, \sigma_i^2)$, i=0,1,2 $\sigma_i^2 \sim Inv-Gamma(0.001, 0.001)$	Combined data set, including both current experiment results and previous ccontrol group experiment

			results

Table 7: Posterior quantities for the regression coefficients and RIp parameters

parameter		Frequentist	Bayesian 1[1]	Bayesian 2[2]	Bayesian 3[3]	Bayesian 4[4]
RI_{25}	PE[a]	1.11	1.07	1.06	0.75	0.75
	SE[b]	0.05	0.09	0.10	0.06	0.06
	95%CI[c]	[1.01,1.22]	[0.87,1.24]	[0.86,1.23]	[0.61,0.86]	[0.62,0.86]
RI_{50}	PE[a]	1.45	1.43	1.42	1.21	1.21
	SE[b]	0.05	0.07	0.07	0.05	0.05
	95%CI[c]	[1.34,1.55]	[1.29,1.56]	[1.29,1.56]	[1.10,1.31]	[1.10,1.31]
β_0	PE[a]	3.16	3.17	3.17	3.37	3.37
	SE[b]	0.04	0.04	0.04	0.01	0.01
	95%CI[c]	[3.08,3.24]	[3.09,3.26]	[3.10,3.26]	[3.35,3.38]	[3.35,3.38]

β_1	PE[a]	0.47	0.4	0.38	-0.07	-0.07
	SE[b]	0.16	0.17	0.17	0.09	0.1
	95%CI[c]	[0.17,0.78]	[0.04,0.71]	[0.04,0.73]	[-0.29,0.08]	[-0.29,0.08]
β_2	PE[a]	-0.66	-0.61	-0.61	-0.41	-0.42
	SE[b]	0.09	0.1	0.11	0.07	0.07
	95%CI[c]	[-0.84,-0.48]	[-0.81,-0.41]	[-0.82,-0.41]	[-0.54,-0.27]	[-0.53,-0.27]

[a]PE= point estimate

[b]SE= standard error of the point estimate

[c]95%CI=the 95% interval estimate

[1]Bayesian 1 = flat prior used for β_0

[2]Bayesian 2 = prior of β_0 generated from summary information

[3]Bayesian 3 = prior of β_0 generated from actual historical observations

[4]Bayesian 4 = flat prior used for β_0, and actual historical control group observations used as part of data

745 **Figures**

746 Figure 1: Histograms of the total young in the control group from the experimental results of lab
747 "CASQS" selected from the data set analyzed in [1] (top = histogram of the experimental results
748 from the historical control group; bottom =Histogram of "current" experiment results produced
749 on May 23, 1992.)

750

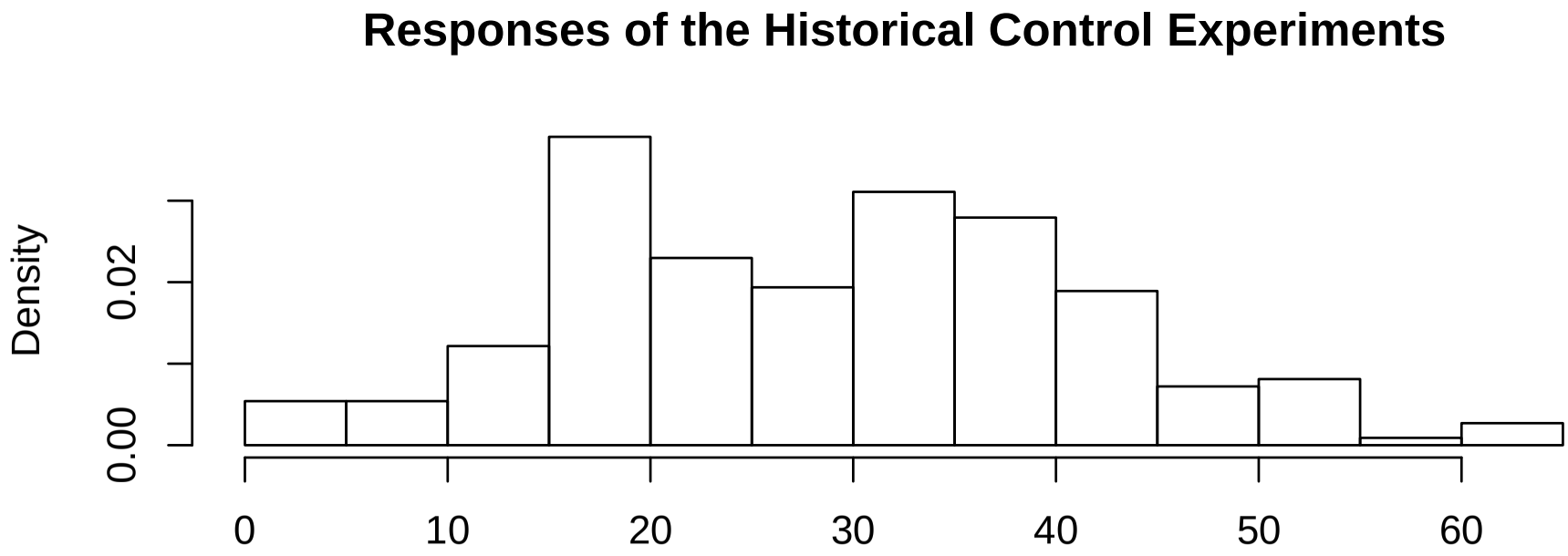
751 Figure 2: Histograms of the RI_{25} and RI_{50} estimates of the 1000 simulated experiments ((a)
752 frequentist estimates of RI_{25}; (b) Bayesian estimates of RI_{25} produced by the Bayesian model
753 using flat priors for all regression coefficients with prior mean 0's; (c) Bayesian estimates of
754 RI_{25} produced by the Bayesian model using flat priors for all regression coefficients with prior
755 mean $\beta_0^0 = \ln(20)$ and $\beta_1^0 = \beta_2^0 = 0$; (d) Bayesian estimates of RI_{25} produced by the Bayesian
756 model using flat priors for all regression coefficients with prior mean $\beta_0^0 = \ln(30)$ and $\beta_1^0 =$
757 $\beta_2^0 = 0$; (e) Bayesian estimates of RI_{25} produced by the Bayesian model using flat priors for
758 $\beta_1$and β_2 with prior mean $\beta_1^0 = \beta_2^0 = 0$ and an informative prior for β_0: $\beta_0 \sim N(\ln(30), 0.001)$;
759 (f) frequentist estimates of RI_{50}; (g) Bayesian estimates of RI_{50} produced by the Bayesian model
760 using flat priors for all regression coefficients with prior mean 0's; (h) Bayesian estimates of
761 RI_{50} produced by the Bayesian model using flat priors for all regression coefficients with prior
762 mean $\beta_0^0 = \ln(20)$ and $\beta_1^0 = \beta_2^0 = 0$; (i) Bayesian estimates of RI_{50} produced by the Bayesian
763 model using flat priors for all regression coefficients with prior mean $\beta_0^0 = \ln(30)$ and $\beta_1^0 =$
764 $\beta_2^0 = 0$; (j) Bayesian estimates of RI_{50} produced by the Bayesian model using flat priors for
765 $\beta_1$and β_2 with prior mean $\beta_1^0 = \beta_2^0 = 0$ and an informative prior for β_0: $\beta_0 \sim N(\ln(30), 0.001)$).

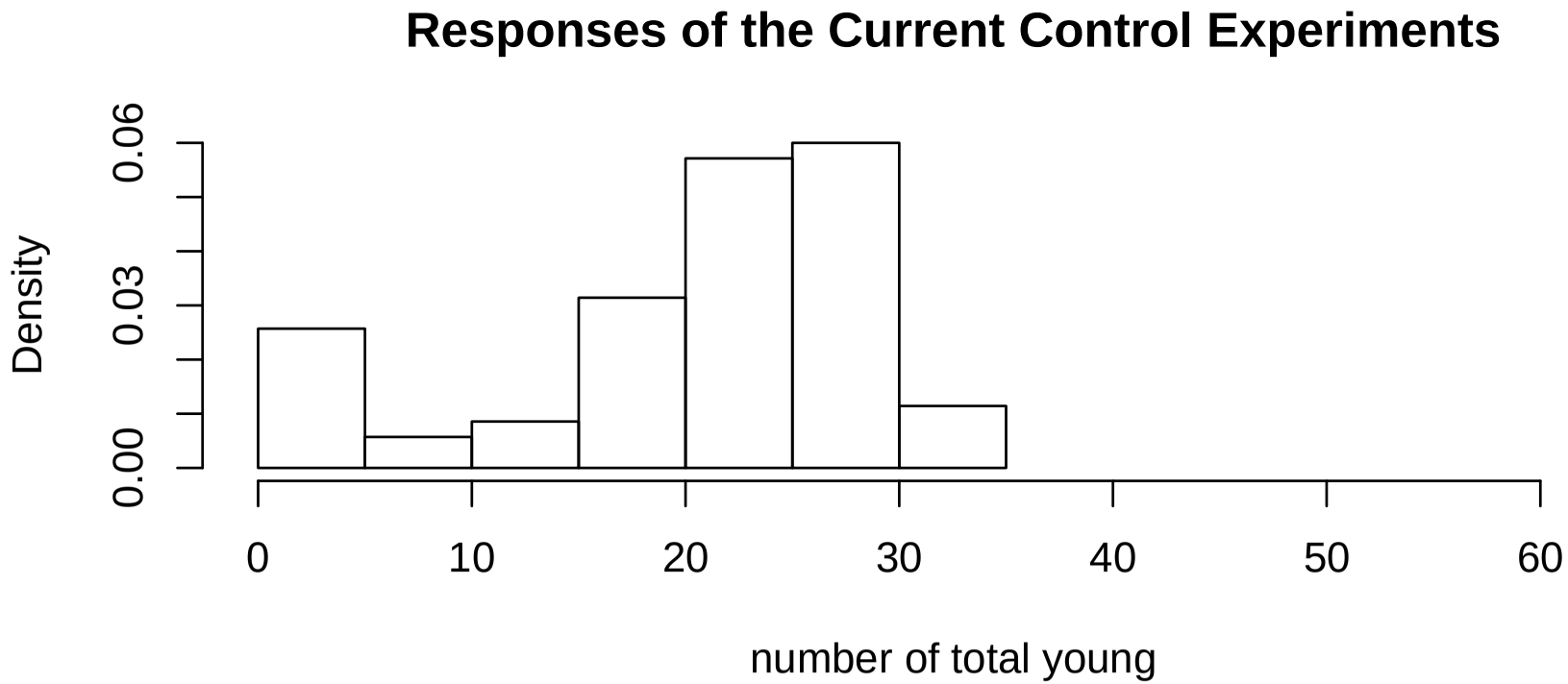
Figure 3: Density curves of the RI_{25} estimates of the 1000 simulated experiments (long-dashed line: frequentist estimates; solid line: Bayesian estimates produced by the Bayesian model using flat priors for all regression coefficients with prior mean 0's; large-dotted line: Bayesian estimates produced by the Bayesian model using flat priors for all regression coefficients with prior mean $\beta_0^0 = \ln(20)$ and $\beta_1^0 = \beta_2^0 = 0$; dotted line: Bayesian estimates produced by the Bayesian model using flat priors for all regression coefficients with prior mean $\beta_0^0 = \ln(30)$ and $\beta_1^0 = \beta_2^0 = 0$; dashed line: Bayesian estimates produced by the Bayesian model using flat priors for $\beta_1$and β_2 with prior mean $\beta_1^0 = \beta_2^0 = 0$ and an informative prior for β_0: $\beta_0 \sim N(\ln(30), 0.001)$).

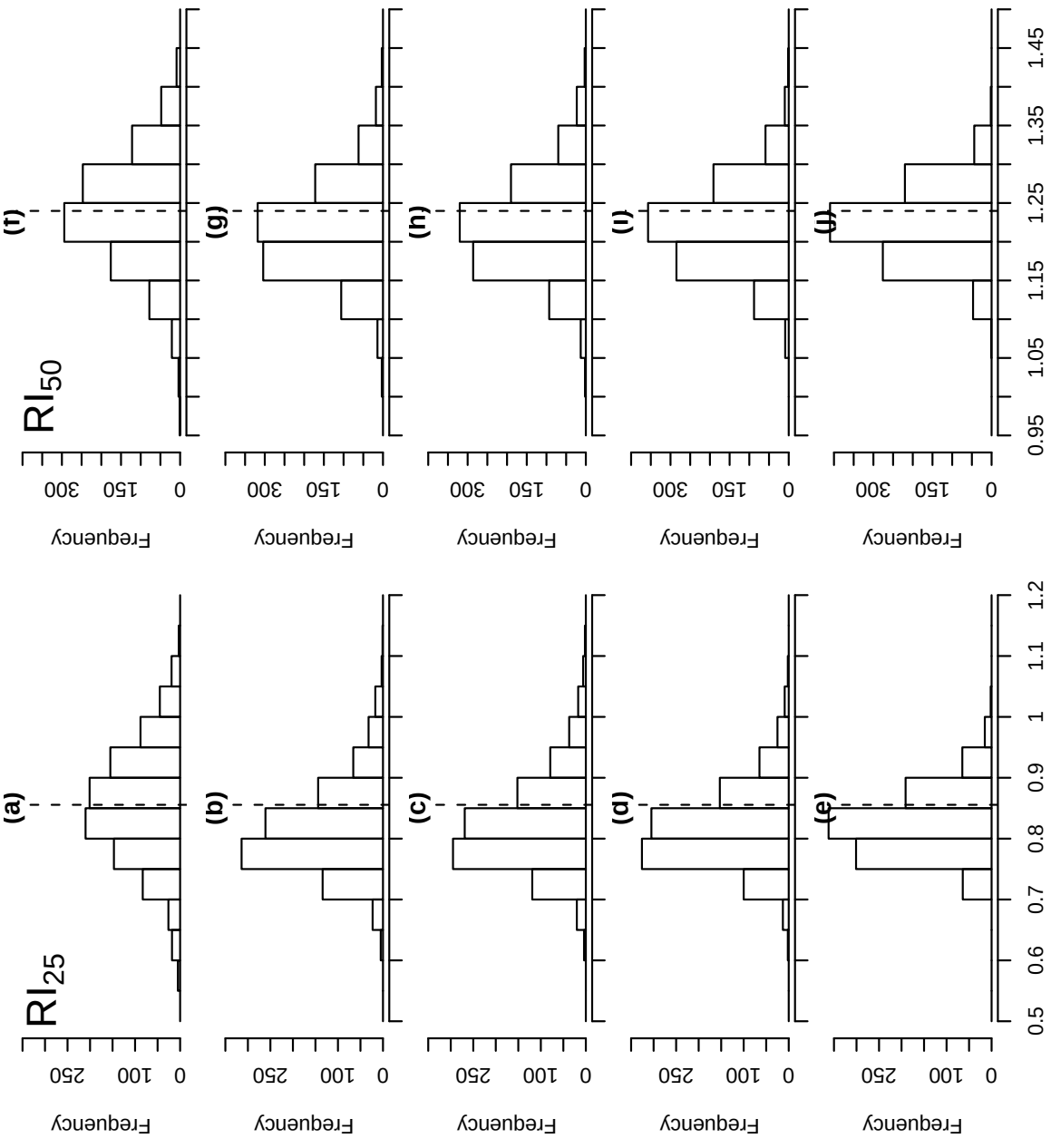
Figure 4: Density curves of the RI_{50} estimates of the 1000 simulated experiments (long-dashed line: frequentist estimates; solid line: Bayesian estimates produced by the Bayesian model using flat priors for all regression coefficients with prior mean 0's; large-dotted line: Bayesian estimates produced by the Bayesian model using flat priors for all regression coefficients with prior mean $\beta_0^0 = \ln(20)$ and $\beta_1^0 = \beta_2^0 = 0$; dotted line: Bayesian estimates produced by the Bayesian model using flat priors for all regression coefficients with prior mean $\beta_0^0 = \ln(30)$ and $\beta_1^0 = \beta_2^0 = 0$; dashed line: Bayesian estimates produced by the Bayesian model using flat priors for $\beta_1$and β_2 with prior mean $\beta_1^0 = \beta_2^0 = 0$ and an informative prior for β_0: $\beta_0 \sim N(\ln(30), 0.001)$).

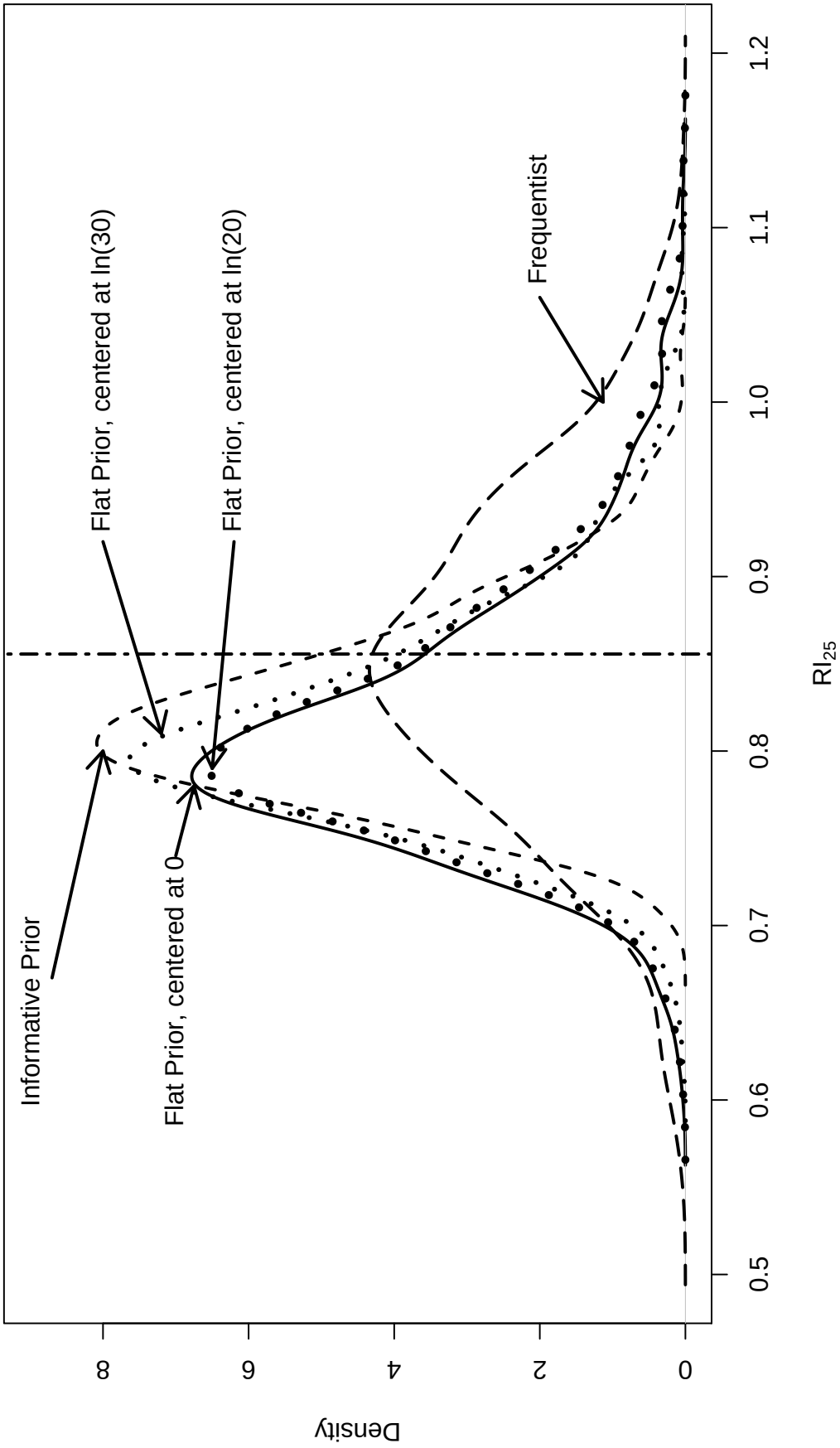
Figure 5: Prior distribution of β_0 generated given the actual historical control data, obtained via the posterior samples produced by the model in Equations 7 and 8.

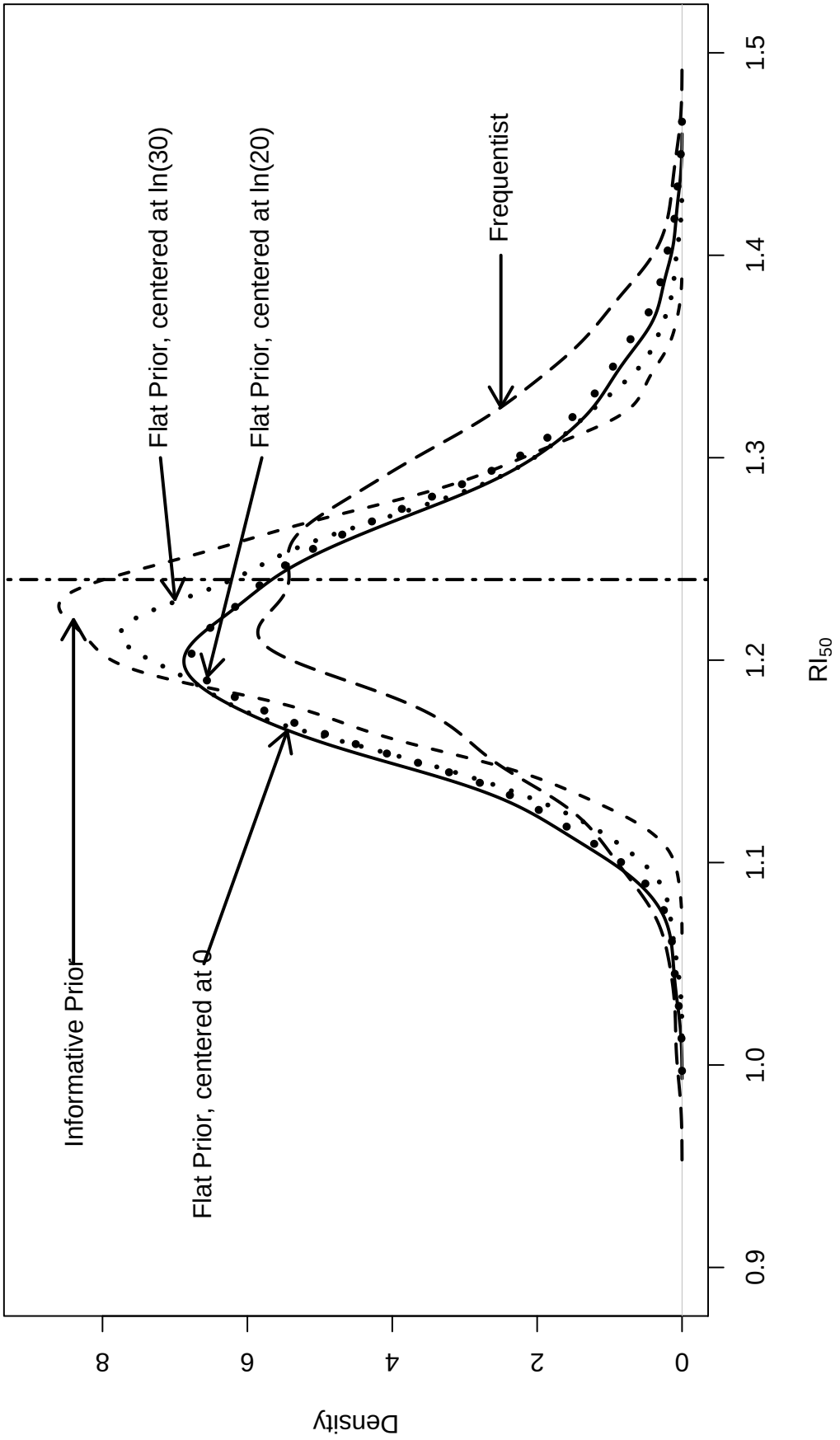
Figure 6: Scatter plot of the total young observed in each concentration group (circles), with the estimated reproduction-concentration curves given by the frequentist (dashed line), Bayesian methods using 3 different priors for β_0 (flat prior and prior generated using summary historical control information: dot-dashed line; prior generated from actual historical observations: long-dashed line) and Bayesian method using the combined data (long-dashed line).

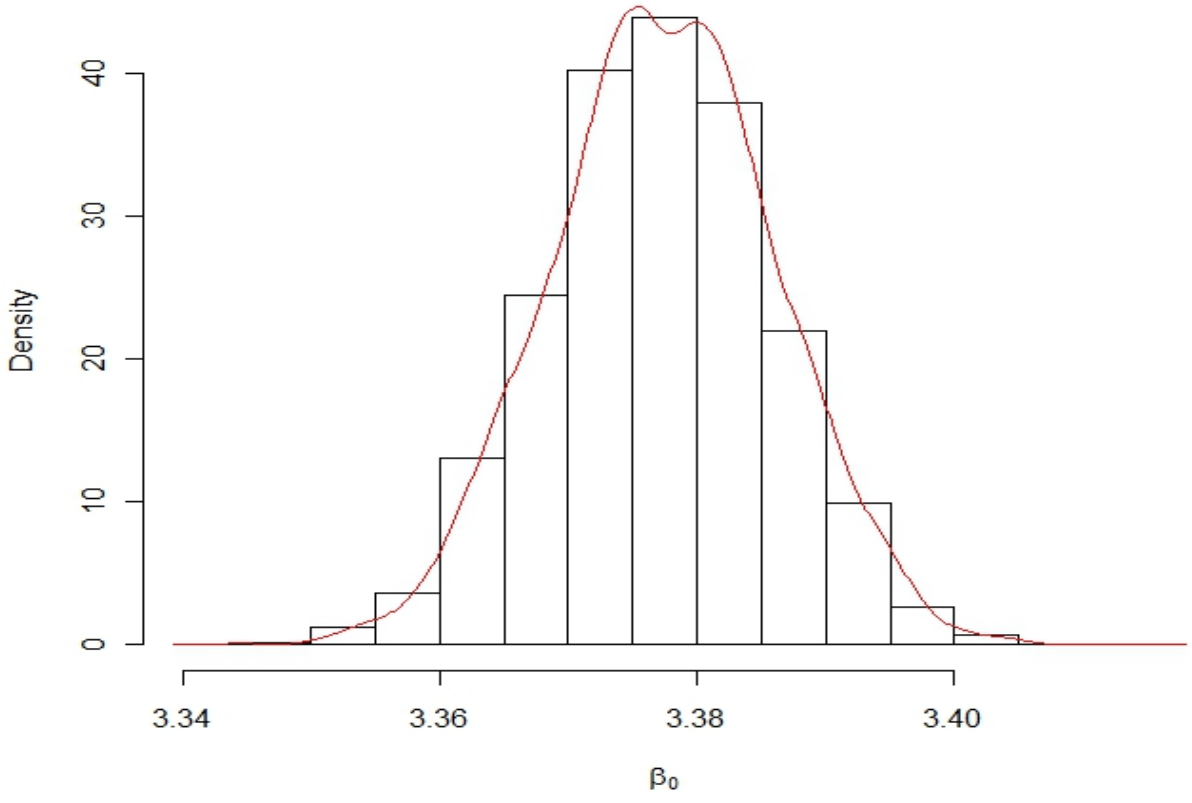

Responses of the Historical Control Experiments
Density
0.00
0.02
0
10
20
30
40
50
60

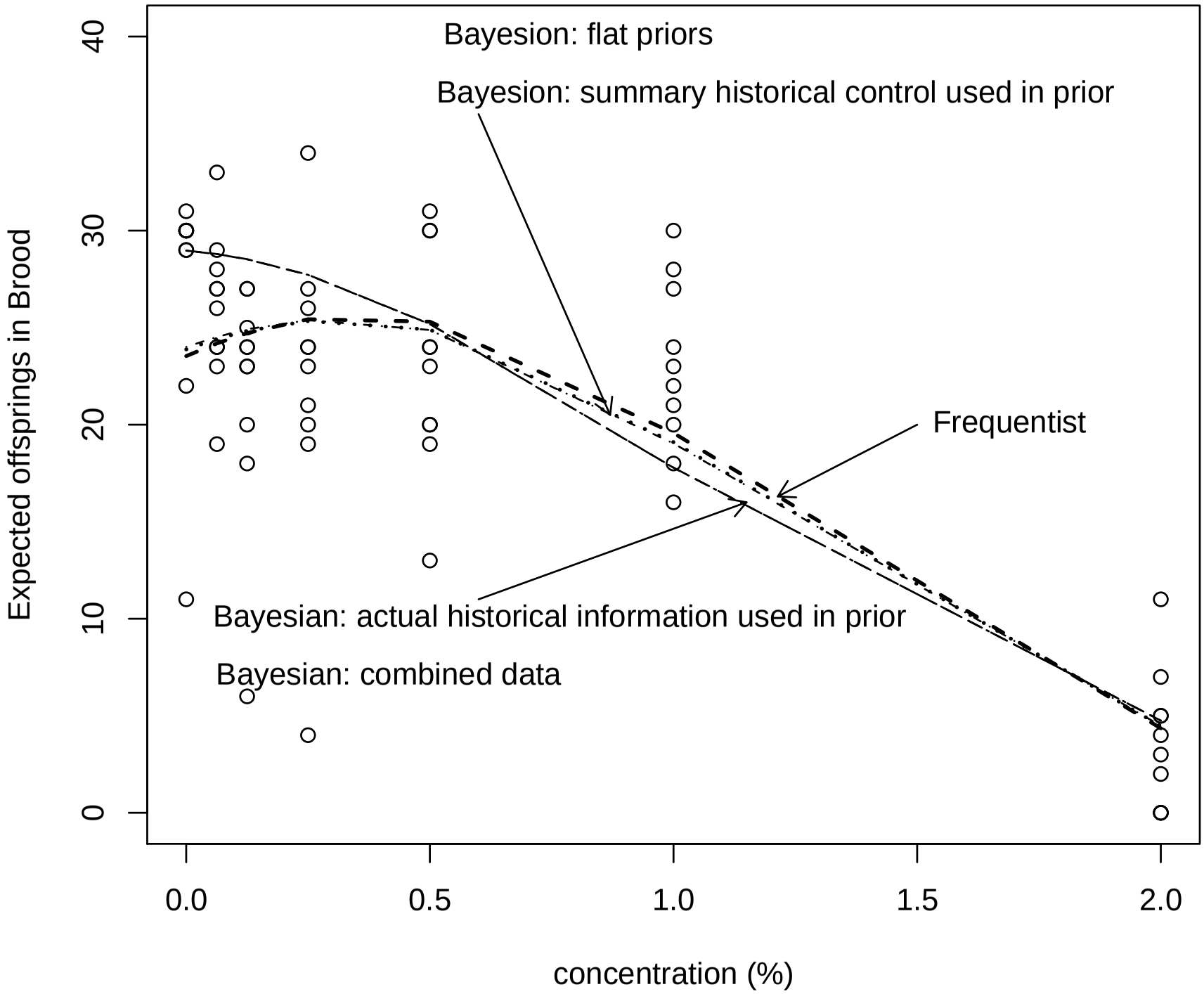

Responses of the Current Control Experiments
Density
0.00
0.03
0.06
0
10
20
30
40
50
60
number of total young


RI25
RI50
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
Frequency
0
100
250
150
300
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
0.95
1.05
1.15
1.25
1.35
1.45


Density
0
2
4
6
8
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
RI25
Informative Prior
Flat Prior, centered at 0
Flat Prior, centered at ln(20)
Flat Prior, centered at ln(30)
Frequentist


Density
0
2
4
6
8
0.9
1.0
1.1
1.2
1.3
1.4
1.5
RI50
Informative Prior
Flat Prior, centered at 0
Flat Prior, centered at ln(20)
Flat Prior, centered at ln(30)
Frequentist


Density
0
10
20
30
40
3.34
3.36
3.38
3.40
β0


Bayesion: flat priors
Bayesion: summary historical control used in prior
Frequentist
Bayesian: actual historical information used in prior
Bayesian: combined data
Expected offsprings in Brood
0
10
20
30
40
0.0
0.5
1.0
1.5
2.0
concentration (%)