Large-Sample Bayesian Posterior Distributions for Probabilistic Sensitivity Analysis

Gordon B. Hazen

Min Huang

IEMS Department

Northwestern University

August 2004
Abstract

In probabilistic sensitivity analyses, analysts assign probability distributions to uncertain model parameters, and use Monte Carlo simulation to estimate the sensitivity of model results to parameter uncertainty. Bayesian methods provide convenient means to obtain probability distributions on parameters given data. We present simple methods for constructing large-sample approximate Bayesian posterior distributions for probabilities, rates, and relative effect parameters, for both controlled and uncontrolled studies, and discuss how to use these posterior distributions in a probabilistic sensitivity analysis. These results draw on and extend procedures from the literature on large-sample Bayesian posterior distributions and Bayesian random effects meta-analysis. We apply these methods to conduct a probabilistic sensitivity analyses for a recently published analysis of zidovudine prophylaxis following rapid HIV testing in labor to prevent vertical HIV transmission in pregnant women. Key Words: decision analysis, cost-effectiveness analysis, probabilistic sensitivity analysis, Bayesian methods, random effects meta analysis, expected value of perfect information, HIV transmission, zidovudine prophylaxis.
Introduction

Sensitivity analysis is today a crucial element in any practical decision analysis, and can play any of several roles in the decision analysis process. Analysts have long recognized the dimensionality limitations of graphically based sensitivity analysis in portraying the robustness of a decision analysis to variations in underlying parameter estimates. If graphical methods allow at most 2- or 3-way sensitivity analyses, how can one be sure that a decision analysis is robust to the simultaneous variation of 10 to 20 parameters?

Probabilistic sensitivity analysis was introduced to address this issue\(^1\). In a probabilistic sensitivity analysis, the analyst assigns distributions to uncertain parameters and can thereby compute as a measure of robustness the probability of a change in the optimal alternative due to variation in an arbitrary set of parameters, or alternately\(^3\), the expected value of perfect information regarding any set of parameters. This computation is most frequently done via Monte Carlo simulation.

The task of fitting distributions to uncertain parameters prior to a probabilistic sensitivity analysis has been approached in several standard ways. Traditionally, distributions of unobservable parameters (such as probabilities or rates) are fitted with a combination of mean and confidence interval estimated from data. Distributions typically used are the beta distribution\(^2\), the logistic-normal distribution\(^1,7\), the uniform distribution and the normal distribution\(^10\). Analysts also use bootstrap methods to obtain sampling distributions\(^11\). For observable parameters (such as costs), these methods are also applicable. However, in this case, it may be more convenient to simply fit a theoretical distribution to the empirical distribution of observations. For example, in Goodman et
Bayesian Posteriors for Probabilistic Sensitivity Analysis

al.\textsuperscript{10}, observations show a peaked distribution, and a triangular distribution was used. Lord, J. et al.\textsuperscript{11} use a piecewise linear approximation to the empirical distribution. Here we will be concerned with obtaining parameter distributions using Bayesian methods. Bayesian methods automatically yield posterior distributions for parameters given observations without any need for distribution fitting, and therefore seem the natural choice for selecting parameter distributions for probabilistic sensitivity analysis. They have not been used extensively for this purpose, however, due to the burden of computing posterior distributions and the inconvenience of specifying the prior distribution required for the Bayesian approach. However, both of these difficulties disappear when the number of observations is large. In this case, the prior distribution has little effect on the resulting posterior, and approximate large-sample normal posterior distributions can be inferred without extensive computation. In the following, we will develop a convenient methodology to obtain large-sample approximate Bayesian posterior distributions and use these distributions in probabilistic sensitivity analyses for probabilities, rates and relative effect parameters. As an example, we apply these methods to conduct a probabilistic sensitivity analysis for a recently published analysis by Mrus and Tsevat of zidovudine prophylaxis following rapid HIV testing in labor to prevent vertical HIV transmission in pregnant women\textsuperscript{12}. This probabilistic sensitivity analysis yields zidovudine prophylaxis optimal 95.9% of the time, and the expected value of perfect information on all relative effect and probability parameters is equal to approximately $10.65 per pregnancy. These results concur with Mrus and Tsevat’s conclusion that the choice of rapid HIV testing followed by zidovudine prophylaxis is not a close call.
Ades, Lu and Claxton\textsuperscript{13} have presented similar methods for computing approximate expected values of sample information, a topic we do not address here. The methods we discuss here overlap with Ades et al. in part, but differ in the use of random effects models for combining heterogeneous studies, where Ades et al. calculate a point estimate for the overall population mean, but we obtain an approximate normal posterior distribution. The latter more accurately reflects population-wide parameter variation. Our material also augments Ades et al. by addressing the issue of combining data from controlled and uncontrolled studies. More generally, the methods we present for obtaining approximate normal posteriors for (possibly transformed) parameters are standard in the Bayesian literature and have been summarized in many texts\textsuperscript{14,15}. However, the specific result we present for the relative efficacy situation in a random effects context has not, to our knowledge, appeared previously.

\textbf{Overview of Model Types}

The influence diagram Figure 1(a) shows the common situation in which observed data $y_i$ from each of $n$ studies $i = 1, \ldots, n$ are influenced by an unknown parameter $\xi$. Typically, $\xi$ is a probability or rate, and $y_i$ is a count of observed critical events in some subject population $i$. The observations $y_1, \ldots, y_n$ inform the choice of a decision or policy whose cost or utility for an individual or group is influenced by the not-yet-observed count $y$ of critical events for that patient or group. The not-yet-observed count $y$ is also influenced by the unknown $\xi$. Analysts can use the observations $y_1, \ldots, y_n$ to make statistical inferences about the unknown $\xi$, and can use these inferences to make predictions about the critical count $y$ and to make recommendations concerning the optimal decision or policy. We will be interested in Bayesian procedures in which the analyst calculates a
posterior probability distribution for $\xi$ and uses it to form a predictive distribution for $y$, from which an expected-utility maximizing decision or policy can be computed.

Many studies compare a treatment with a control, and for these, the underlying parameter $\xi$ must be taken to be a vector $\xi = (\xi^0, \epsilon)$, where $\xi^0$ is the unknown probability or rate for the control group and $\epsilon$ is the unknown efficacy, some measure of the effectiveness of the treatment. The unknown probability or rate for the treatment group may be $\xi^1 = \xi^0 + \epsilon$, or $\xi^1 = \xi^0 \cdot \epsilon$, or some other function of $\xi^0$ and $\epsilon$, depending on model specifics. In this situation, the observations $y_i$ are also vectors $y_i = (y_i^0, y_i^1)$, or perhaps $y_i = (y_i^0, y_i^1 - y_i^0)$, one component for control and one for treatment or treatment effect, and the decision or policy in question might be whether the treatment is cost-effective, or beneficial in terms of expected utility.

Figure 1(a) assumes that the unknown parameter $\xi$ has the same value for each population $i$, that is, the populations are homogenous. One may test this assumption statistically, as we discuss below. If homogeneity is confirmed, it is common to proceed with analyses that effectively pool the data as though they are from the same source, as shown in Figure 1(b).
Figure 1: (a) An influence diagram depicting the cost-effectiveness or decision-analytic setting when data from homogenous studies is available to estimate a parameter $\xi$. In a Bayesian approach, the posterior distribution on $\xi$ given $y_1, \ldots, y_n$ can be used to determine the optimal decision or policy given the observations $y_1, \ldots, y_n$. In this situation, observations are typically pooled as if they come from a single study, as shown in (b).

If homogeneity fails, then the populations are heterogeneous, with a different value $\xi_i$ of the unknown parameter for each population $i$. This situation is depicted in Figure 2, which shows a random effects model of heterogeneity, in which it is assumed that the values $\xi_i$ are independent draws from a population of $\xi$-values with mean $\mu$. Once again the observations $y_1, \ldots, y_n$ inform the choice of a decision or policy whose cost or utility is influenced by the not-yet-observed count $y$ of critical events for a patient or group.

However, now $y$ is influenced by an unknown $\xi$ drawn from the same population of $\xi$-values with mean $\mu$. The analyst aims to use the observations $y_1, \ldots, y_n$ to make inferences about $\mu$ and in turn about $\xi$, from which predictions of the critical count $y$ can be made. A Bayesian analyst would seek a posterior distribution on $\mu$, from which could be
Bayesian Posteriors for Probabilistic Sensitivity Analysis

inferred a predictive distribution on $\xi$. This in turn would allow the computation of a utility-maximizing decision or policy given the observations $y_1, \ldots, y_n$. This situation also allows for controlled studies. Again the observations $y_i = (y_i^0, y_i^1)$ or perhaps $y_i = (y_i^0, y_i^1 - y_i^0)$ would be for control and treatment groups, the unknown $\xi_i = (\xi_i^0, \varepsilon_i)$ would consist of a component $\xi_i^0$ for the $i^{th}$ control group, and an efficacy parameter $\varepsilon_i$ for the $i^{th}$ group, and $\mu = (\mu^0, \mu^1)$ would also have components for control and efficacy.

Figure 2: An influence diagram depicting the cost-effectiveness or decision-analytic setting when data from heterogeneous studies is available to estimate a parameter $\xi$. Past studies aimed at estimating $\xi$ are portrayed using a random effects model in which the corresponding parameter $\xi_i$ for the $i^{th}$ study is sampled from a population with mean $\mu$. In a Bayesian approach, the posterior distribution on $\mu$ given $y_1, \ldots, y_n$ is used to infer a predictive distribution on $\xi$. This predictive distribution can be used to determine the optimal decision or policy given the observations $y_1, \ldots, y_n$.

In the sequel, we examine in turn each combination of these two dichotomies (homogenous/heterogeneous, uncontrolled/controlled), summarizing how to obtain large-sample approximate Bayesian posterior distributions for $\xi$ (in the homogeneous case) or
Bayesian Posteriors for Probabilistic Sensitivity Analysis

149  \( \mu \) (in the heterogeneous case), and how to use these posterior distributions to conduct a probabilistic sensitivity analysis using Monte Carlo simulation. In each situation we illustrate our results with data cited by Mrus and Tsevat\(^{12} \) in their analysis of zidovudine prophylaxis following rapid HIV testing in labor. We conclude, as mentioned above, with a complete probabilistic sensitivity analysis of all probability and efficacy parameters in the Mrus and Tsevat model.

155  **Observations from a Single Study**

156  If, as in Figure 1, there is only a single study, or if there are multiple studies that pass a homogeneity test and may be pooled, then the posterior distribution of the unknown parameter \( \xi \) can be calculated exactly using standard conjugate Bayesian methods. These are summarized in Table 1.1. Exact posterior distributions may be obtained when \( \xi \) is a probability \( p \), a rate \( \lambda \), or a mean. The beta, gamma and normal distributions mentioned may be found in standard probability texts. The table also illustrates how posterior distributions simplify when the prior is noninformative, reflecting absence of any prior data.
### Table 1.1: Single-study models using conjugate prior distributions

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Observations</th>
<th>Prior distributions</th>
<th>Posterior distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability $p$</td>
<td>$k$ events in $n$ independent trials; $k \sim$ binomial($n$, $p$)</td>
<td>$p \sim$ beta($\alpha$, $\beta$)</td>
<td>$p \sim$ beta($\alpha + k$, $\beta + n - k$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noninformative uniform prior on $p$ ($\alpha = 1$, $\beta = 1$)</td>
<td>$p \sim$ beta($k + 1$, $n - k + 1$)</td>
</tr>
<tr>
<td>Rate $\lambda$</td>
<td>$k$ events in duration $\Delta t$; $k \sim$ Poisson($\lambda \Delta t$)</td>
<td>$\lambda \sim$ gamma($r$, $\theta$)</td>
<td>$\lambda \sim$ gamma($r + k$, $\theta + \Delta t$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Noninformative prior on $\lambda$ ($r = 0$, $\theta = 0$)</td>
<td>$\lambda \sim$ gamma($k$, $\Delta t$)</td>
</tr>
<tr>
<td>Mean $\xi$</td>
<td>$y \sim$ normal($\xi$, $\sigma^2$)</td>
<td>$\xi \sim$ normal($\xi_0$, $\sigma_0^2$)</td>
<td>$\xi \sim$ normal($\xi$, $\sigma^2$)</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$ known or estimated</td>
<td>$\xi^2 \sim$ normal($\xi_0$, $\sigma_0^2$)</td>
<td>$\xi_0 = \frac{\sigma_0^2 \xi + \sigma^2 y}{\sigma_0^2 + \sigma^2}$, $\sigma_0^2 = \sigma^2 + \sigma_0^{-2}$</td>
</tr>
<tr>
<td></td>
<td>noninformative uniform prior on $\xi$ ($\sigma_0^{-2} = 0$)</td>
<td>$\xi \sim$ normal($y$, $\sigma^2$)</td>
<td></td>
</tr>
</tbody>
</table>

If sample size is large, and parameters are appropriately transformed, then posterior distributions may be approximated by normal distributions in the manner depicted in Table 1.2. For probability parameters $p$, the appropriate transformation is the log transformation $\xi = \ln p$, or the logit transformation $\xi = \text{logit}(p) = \ln (p/(1-p))$; and for rate parameters $\lambda$, the log transformation $\xi = \ln \lambda$ is appropriate. The observations $k$ are transformed similarly to approximately normal observations $y$. After transformation, the conjugate normal posterior from Table 1.1 can be applied. The large sample size implies that $\sigma^{-2}$ is very large compared to $\sigma_0^{-2}$, which may be effectively assumed zero, so the noninformative case from Table 1.1 applies.

The final column of Table 1.2 indicates how a random parameter value may be generated for Monte Carlo simulation: First generate a random $\xi$ from its approximate posterior normal($y$, $\sigma^2$) distribution, then transform $\xi$ back to obtain a random value of the original
parameter $p$ or $\lambda$. These large-sample approximate normal posteriors are unnecessary for this simple case, as the conjugate beta or gamma posteriors from Table 1.1 are exactly correct and easily computed. However, the large-sample approximations generalize to more complicated situations that we will discuss below, where conjugate prior models are no longer tractable. We present them here to illustrate our general approach in a simple situation.

Table 1.2: Large-sample approximate single study models for probabilities and rates.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Observations</th>
<th>Transformation to an approximate normal model (Table 1.1)</th>
<th>Generating new parameter values for MC simulation: First generate $\xi \sim \text{normal}(y, \sigma^2)$, then:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability $p$</td>
<td>$k$ events in $n$ independent trials $k \sim \text{binomial}(n, p)$</td>
<td>$\xi = \text{logit}(p)$</td>
<td>Calculate $p = \frac{e^\xi}{1 + e^\xi}$.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$y = \text{logit}(k/n)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2 = 1/k + 1/(n-k)$</td>
<td></td>
</tr>
<tr>
<td>Rate $\lambda$</td>
<td>$k$ events in duration $\Delta t$ $k \sim \text{Poisson}(\lambda \Delta t)$</td>
<td>$\xi = \log(\lambda)$</td>
<td>Calculate $\lambda = e^\xi$.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$y = \log(k/\Delta t)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2 = 1/k$</td>
<td></td>
</tr>
</tbody>
</table>

**Example 1: Acceptance rate for rapid HIV testing and treatment**

Mrus and Tsevat cite Rajegowda et al\textsuperscript{16}, who observed that in $n = 539$ patients, $k = 462$ (85.71\%) were willing to accept rapid HIV test and treatment. To obtain a Bayesian posterior on the acceptance rate $p$, we may use the conjugate beta posterior in Table 1.1 or either of the large sample approximate posteriors for binomial data in Table 1.2. The former yields a beta$(k+1, n-k+1) = \text{beta}(463, 78)$ posterior. The normal log-odds model yields $y = \text{logit}(k/n) = 1.7918$ and $\sigma = (1/k + 1/(n-k))^{1/2} = 0.1231$ and hence a normal posterior with mean 1.7918 and standard deviation 0.1231 on $\xi = \text{logit}(p)$. The lognormal
model yields $y = \log(k/n) = -0.1542$ and $\sigma = (1/k - 1/n)^{1/2} = 0.0176$ and hence a normal posterior with mean $-0.1542$ and standard deviation $0.0176$ on $\xi = \log(p)$. (Here all logarithms are to the natural base $e$.) The beta posterior and the implied normal log-odds and lognormal posteriors on $p$ are graphed in Figure 3, and are virtually identical, reflecting the fidelity of the large-sample normal approximation. To generate a random value of $p$ for the purposes of Monte Carlo simulation, we may sample directly from the beta(463,78) posterior. Alternately, as indicated in Table 1.2, we may sample $\xi$ from either of the normal posteriors just derived, and then transform via $p = e^{\xi}/(1+e^{\xi})$ or $p = e^{\xi}$ as appropriate to obtain a random $p$.  

![Graph showing beta, normal log-odds, and lognormal approximate posterior distributions for the acceptance rate $p$ for HIV testing and treatment, based on data from Rajegowda et al. The approximate normal log-odds and lognormal posteriors are so close to the true beta posterior that it is difficult to distinguish them.]

**Observations From a Single Controlled Study**

We consider now the case of observations from a single controlled study, or from several controlled studies that may be pooled. Table 2.1 shows the situation in which the...
observations \( y_0, y_1 \) from the control and treatment groups are normally distributed. Here

the parameter vector \( (\xi^0, \varepsilon) \) consists of the unknown mean \( \xi^0 \) in the control group and the unknown efficacy \( \varepsilon \). The mean in the treatment group is then \( \xi^1 = \xi^0 + \varepsilon \). Table 2.1 gives the posterior distribution of \( (\xi^0, \varepsilon) \) under a noninformative prior when all likelihoods are normal. Note that \( \xi^0 \) and \( \varepsilon \) are dependent \textit{a posteriori}, and the table gives the marginal distribution of \( \xi^0 \) and the conditional distribution of \( \varepsilon \) given \( \xi^0 \). Generating parameter values \( \xi^0 \) and \( \xi^1 \) for Monte Carlo simulation is then an easy 3-step process, as summarized in the table.

**Table 2.1: Normally distributed observations from a single controlled study.**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Observations</th>
<th>Prior distributions</th>
<th>Posterior distribution</th>
<th>Generating new parameter values ( \xi^0 ), ( \xi^1 ) for Monte Carlo simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi^0, \xi^1 )</td>
<td>( y_0 \sim \text{normal}(\xi^0, \sigma^2_0) )</td>
<td>( \xi^0, \varepsilon ) are independent, each with a noninformative prior</td>
<td>( \xi^0 \sim \text{normal}(y_0, \sigma^2_0) )</td>
<td>Step 1. Generate ( \xi^0 \sim \text{normal}(y_0, \sigma^2_0) ).</td>
</tr>
<tr>
<td>( \varepsilon = \xi^1 - \xi^0 )</td>
<td>( y_1 \sim \text{normal}(\xi^1, \sigma^2_1) )</td>
<td></td>
<td>( \varepsilon \mid \xi^0 \sim \text{normal}(y_1 - \xi^0, \sigma^2_1) )</td>
<td>Step 2. Generate ( \varepsilon \sim \text{normal}(y_1 - \xi^0, \sigma^2_1) ).</td>
</tr>
<tr>
<td>( \sigma^2_0, \sigma^2_1 ) known or estimated from data.</td>
<td></td>
<td></td>
<td></td>
<td>Step 3. Calculate ( \xi^1 = \xi^0 + \varepsilon ).</td>
</tr>
</tbody>
</table>

Unlike the case of a single uncontrolled study, controlled studies involving binomial or Poisson observations do not have tractable conjugate Bayesian updates. However, for suitable transformations of parameters and observations summarized in Table 2.2, large-sample normal approximations are valid, and results from Table 2.1 may be applied. Table 2.2 gives the resulting approximate normal posterior distributions. To generate random variates for a probabilistic sensitivity analysis, one may use the procedure from
Bayesian Posteriors for Probabilistic Sensitivity Analysis

Table 2.1 and then transform back as indicated in Table 2.2 to obtain the desired probability or rate random variates.

Table 2.2: Controlled studies involving binomial or Poisson observations and their large-sample normal approximation.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Observations</th>
<th>Transformation to an approximate normal model (Table 2.1)</th>
<th>Generating a new parameter value for MC simulation: Steps 1, 2 and 3 are as in Table 2.1. Then</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities $p_0, p_1$</td>
<td>$k_0$ events in $n_0$ independent trials without intervention $k_1$ events in $n_1$ independent trials with intervention $k_0 \sim \text{binomial}(n_0, p_0)$ $k_1 \sim \text{binomial}(n_1, p_1)$</td>
<td>$\xi_0 = \logit(p_0)$, $\xi_1 = \logit(p_1)$, $y^0 = \log(k_0/n_0)$, $y^1 = \log(k_1/n_1)$, $\sigma_0^2 = 1/k_0 + 1/(n_0 - k_0)$, $\sigma_1^2 = 1/k_1 + 1/(n_1 - k_1)$</td>
<td>Step 4. Calculate $p_0 = \frac{e^{\xi_0}}{1 + e^{\xi_0}}$, $p_1 = \frac{e^{\xi_1}}{1 + e^{\xi_1}}$</td>
</tr>
<tr>
<td>Probabilities $p_0, p_1$</td>
<td>$k_0$ events in $n_0$ independent trials without intervention $k_1$ events in $n_1$ independent trials with intervention $k_0 \sim \text{binomial}(n_0, p_0)$ $k_1 \sim \text{binomial}(n_1, p_1)$</td>
<td>$\xi_0 = \log(p_0)$, $\xi_1 = \log(p_1)$, $y^0 = \log(k_0/n_0)$, $y^1 = \log(k_1/n_1)$, $\sigma_0^2 = 1/k_0 - 1/n_0$, $\sigma_1^2 = 1/k_1 - 1/n_1$</td>
<td>Step 4. Calculate $p_0 = e^{\xi_0}$, $p_1 = e^{\xi_1}$, or $RR = e^{\xi}$</td>
</tr>
<tr>
<td>Rates $\lambda_0, \lambda_1$</td>
<td>$k_0$ events in duration $\Delta t_0$ without intervention $k_1$ events in duration $\Delta t_1$ with intervention $k_0 \sim \text{Poisson}(\lambda_0 \Delta t_0)$ $k_1 \sim \text{Poisson}(\lambda_1 \Delta t_1)$</td>
<td>$\xi_0 = \log(\lambda_0)$, $\xi_1 = \log(\lambda_1)$, $y^0 = \log(k_0/\Delta t_0)$, $y^1 = \log(k_1/\Delta t_1)$, $\sigma_0^2 = 1/k_0$, $\sigma_1^2 = 1/k_1$</td>
<td>Step 4. Calculate $\lambda_0 = e^{\xi_0}$, $\lambda_1 = e^{\xi_1}$, or $RR = e^{\xi}$</td>
</tr>
</tbody>
</table>

Example 2: The effect of zidovudine prophylaxis on HIV transmission

Mrus and Tsevat cite seven controlled studies\textsuperscript{17,18,19,20,21,22,23} giving data for estimating the effect of zidovudine prophylaxis on mother-to-infant HIV transmission risk. Observed

Example 2: The effect of zidovudine prophylaxis on HIV transmission

Mrus and Tsevat cite seven controlled studies\textsuperscript{17,18,19,20,21,22,23} giving data for estimating the effect of zidovudine prophylaxis on mother-to-infant HIV transmission risk. Observed
data are listed as follows, where \( k_i^0 \) of \( n_i^0 \) infants in the control group \( i \) are HIV infected, and \( k_i^1 \) of \( n_i^1 \) infants in the prophylaxis group \( i \) are HIV infected, \( i=1,2,...7 \).

<table>
<thead>
<tr>
<th>Group</th>
<th>Without zidovudine</th>
<th>With zidovudine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( k_i^0 )</td>
<td>( n_i^0 )</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
<td>152</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>216</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>115</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>198</td>
</tr>
<tr>
<td>7</td>
<td>1019</td>
<td>5571</td>
</tr>
</tbody>
</table>

Assuming homogeneity tests are passed (we will discuss this further below), we can pool the 7 study groups into one as follows.

<table>
<thead>
<tr>
<th>Without zidovudine</th>
<th>With zidovudine</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k^0 )</td>
<td>1220</td>
</tr>
<tr>
<td>( n^0 )</td>
<td>6363</td>
</tr>
<tr>
<td>( k^0 / n^0 )</td>
<td>0.191733</td>
</tr>
<tr>
<td>( y^0 = \log(k^0 / n^0) )</td>
<td>-1.65165</td>
</tr>
<tr>
<td>( \sigma_0^2 = 1/k^0 - 1/n^0 )</td>
<td>0.000663</td>
</tr>
</tbody>
</table>

According to Table 2.1, the transformed log-risk parameter \( \xi_0 = \log p_0 \) has a posterior approximate normal distribution with mean \(-1.65165\) and standard deviation \(0.000663^{1/2} = 0.0257\). Moreover, the efficacy parameter \( \varepsilon \) has, given \( \xi_0 \), an approximate normal distribution with mean \(-2.41004 - \xi_0\) and standard deviation \(0.002844^{1/2} = 0.0533\).

**Observations From Several Heterogeneous Studies**

Consider now the random effects situation depicted in Figure 2, in which observations relevant to a parameter \( \xi \) arise from several heterogeneous studies randomly drawn from an overall population with mean \( \mu \). Let \( \sigma_i^2 \) be the variance of this overall population. If
$\sigma_{\xi}^2$ is positive, then the studies are heterogeneous, whereas if $\sigma_{\xi}^2 = 0$, then all $\xi_i$ are equal and the homogenous case of Figure 1 obtains. A popular test $^{24}$ of the null hypothesis $\sigma_{\xi}^2 = 0$ uses the statistic $Q$ given by

$$Q = \sum_{i=1}^{n} \omega_i^{null} (y_i - \bar{y}^{null})^2$$

$$\omega_i^{null} = \frac{1}{\sigma_{ii}^{null}}$$

$$\bar{y}^{null} = \frac{\sum_{i=1}^{n} \omega_i^{null} y_i}{\sum_{i=1}^{n} \omega_i^{null}}.$$  (1)

$Q$ has an approximate $\chi^2$ distribution with $n-1$ degrees of freedom under the null hypothesis, so one would reject this hypothesis if $Q > \chi^2_{\alpha}(n-1)$ for some appropriate test level $\alpha$. However, it is known that this test has low power. If $\sigma_{\xi}^2 = 0$ is rejected, then a moment estimator of $\sigma_{\xi}^2$ is $^{24}$

$$\hat{\sigma}_{\xi}^2 = \frac{(Q - (n-1))^{+}}{\sum_{i=1}^{n} \omega_i^{null} - \left( \sum_{i=1}^{n} (\omega_i^{null})^2 / \sum_{i=1}^{n} \omega_i^{null} \right)}.$$  (2)

Table 3.1 gives the posterior distribution for $\mu$ for heterogeneous studies when all likelihoods are normal, and a two-step Monte Carlo procedure for probabilistic sensitivity analysis. This two-step procedure may be collapsed into a single step, as the posterior distribution of $\xi$ is normal($\bar{y}, \sigma_{\xi}^2 + (\sum \omega_i)^{-1}$) and one may sample directly from this distribution. Table 3.2 gives large-sample approximate posterior distributions for transformed probabilities and rates under binomial and Poisson sampling, drawing on the results in Table 3.1.
Table 3.1: Normally distributed observations from heterogeneous groups (random effects model)

<table>
<thead>
<tr>
<th>Parameter for study $i$</th>
<th>Observations $y_i$ for study $i$</th>
<th>Prior distributions</th>
<th>Posterior distribution</th>
<th>Generating a new parameter value $\xi$ for Monte Carlo simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_i$</td>
<td>$y_i \sim \text{normal}(\xi_i, \sigma_i^2)$</td>
<td>$\mu$ has a noninformative prior $\xi_i</td>
<td>\mu \sim \text{normal}(\bar{y}, (\sum_i \omega_i)^{-1})$</td>
<td>Step 1. Generate $\mu \sim \text{normal}(\bar{y}, (\sum_i \omega_i)^{-1})$.</td>
</tr>
<tr>
<td></td>
<td>$\sigma_i^2$ known or estimated.</td>
<td>$\sigma_i^2 \sim \text{normal}(\mu, \sigma \xi^2)$</td>
<td>where $\omega_i = (\sigma_i^2 + \sigma \xi^2)^{-1}$</td>
<td>Step 2. Generate $\xi \sim \text{normal}(\mu, \sigma \xi^2)$.</td>
</tr>
</tbody>
</table>

Table 3.2: Binomial and Poisson observations from heterogeneous groups and their large-sample normal approximation.

<table>
<thead>
<tr>
<th>Parameter for study $i$</th>
<th>Observations for study $i$</th>
<th>Transformation to an approximate normal random effects model (Table 3.1)</th>
<th>Generating new parameter value for MC simulation: Steps 1 and 2 are as in Table 3.1. Then:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability $p_i$</td>
<td>$k_i$ events in $n_i$ independent trials. $k_i</td>
<td>p_i \sim \text{binomial}(n_i, p_i)$</td>
<td>$\xi_i = \logit(p_i)$, $y_i = \logit(k_i/n_i)$, $\sigma_i^2 = 1/k_i + 1/(n_i-k_i)$.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Step 3. Calculate $p = e^{\xi}$.</td>
</tr>
<tr>
<td>Rate $\lambda_i$</td>
<td>$k_i$ events in duration $\Delta t_i$. $k_i</td>
<td>\lambda_i \sim \text{Poisson}(\lambda_i \Delta t_i)$</td>
<td>$\xi_i = \log(\lambda_i)$, $y_i = \log(k_i/\Delta t_i)$, $\sigma_i^2 = 1/k_i$.</td>
</tr>
</tbody>
</table>

Example 3: Specificity of rapid HIV testing

Mrus and Tsevat cite three studies providing data on the specificity of rapid HIV testing, observing that in $n_i$ people who were not HIV-infected, there were $k_i$ having negative test result, $i = 1,2,3$. The data and its logit transformation are as follows.
Bayesian Posteriors for Probabilistic Sensitivity Analysis

<table>
<thead>
<tr>
<th>Group</th>
<th>$k_i$</th>
<th>$n_i$</th>
<th>$\hat{p}_i = k_i/n_i$</th>
<th>$y_i = \text{logit}(k_i/n_i)$</th>
<th>$\sigma_i^2 = 1/k_i + 1/(n_i-k_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>446</td>
<td>451</td>
<td>0.988914</td>
<td>4.490881</td>
<td>0.202242</td>
</tr>
<tr>
<td>2</td>
<td>783</td>
<td>790</td>
<td>0.991139</td>
<td>4.717223</td>
<td>0.144134</td>
</tr>
<tr>
<td>3</td>
<td>486</td>
<td>496</td>
<td>0.979839</td>
<td>3.883624</td>
<td>0.102058</td>
</tr>
</tbody>
</table>

The test statistic (1) for the null hypothesis $\sigma_\xi^2 = 0$ of homogeneity yields $Q = \sum_{i=1}^{3} \omega_i^{null} (y_i - \bar{y}^{null})^2 = 3.08$, with $p$-value = 0.21, insufficient to reject $\sigma_\xi^2 = 0$. However, due to the low power of this test, we may consider this insufficient evidence for homogeneity and continue to assume heterogeneity. In this case, the moment estimator (2) yields $\hat{\sigma}_\xi^2 = 0.07795$.

Using this estimate and the results in Table 3.1 and Table 3.2 under the logit transformation, we obtain $\left(\sum_i \omega_i^{-1}\right)^{-1} = 0.07338$, $\sum_i \omega_i y_i = 58.84$, $\bar{y} = 4.318$. Hence the approximate posterior distribution of parameter $\mu$ is normal with mean 4.318, and variance 0.07338.

Random variate generation is then a three-stage process, as indicated in Table 3.1 and Table 3.2: First generate $\mu$ from this normal distribution, then generate $\xi$ from a normal distribution with mean $\mu$ and variance $\hat{\sigma}_\xi^2 = 0.07795$, and then transform via $p = \frac{e^\xi}{1+e^\xi}$ to obtain a random value of the specificity $p$. Alternately, we could generate $\xi$ directly from a normal distribution with mean 4.318 and standard deviation $(0.07338 + 0.07795)^{1/2} = 0.389$.

Had we accepted homogeneity as indicated by the $Q$ statistic, we would have pooled the three studies to obtain $k = 1715$ negative test results in $n = 1737$ HIV-negative individuals.

Using the logit transformation from Table 1.2, we obtain $y = \text{logit}(k/n) = 4.356$, $\sigma^2 = 1/k$.
+ \frac{1}{n-k} = 0.04604 and a posterior distribution on $\xi$ that is normal with mean 4.356 and standard deviation $0.04604^{1/2} = 0.215$. This has virtually the same mean as the unpooled case but roughly half the standard deviation.

The implied pooled and unpooled posterior distributions for the specificity $p$ are graphed in Figure 4. The effect of retaining heterogeneity is a wider posterior on $p$ than would be obtained by accepting homogeneity and pooling. Were the homogeneity hypothesis correct and $\sigma_{\xi}^2$ really equal to zero, the two posteriors would be nearly identical† – here the wider posterior for the unpooled case reflects residual uncertainty about homogeneity. So in this light the decision to retain heterogeneity as a possibility seems prudent.

![Figure 4. Log-odds normal posterior distributions on the specificity $p$ of rapid HIV test, one based on pooling three studies under the assumption they are homogeneous, and the other unpooled result based on combining studies using the random effects model. The latter gives a wider posterior distribution, reflecting residual uncertainty regarding whether the studies are really homogeneous.](image)

† They would be exactly the same if instead of pooling studies before applying the normal approximation, as was done here, we pooled after applying a normal approximation to each study.
Observations From Several Heterogeneous Controlled Studies

For heterogeneous controlled studies, each element in Figure 2 is a two-dimensional vector consisting of baseline and efficacy components: \( \mu = (\mu^0, \mu^e)^T, \xi_i = (\xi^0_i, \xi^e_i)^T \), where the superscript \( T \) denotes vector transpose. If \( y^0_i, y^e_i \) are the (possibly transformed) control and treatment observations in study \( i \), we adopt, for consistency, the notation \( y_i = (y^0_i, y^e_i - y^0_i)^T \), which records the control observation and the treatment effect \( y^e_i - y^0_i \).

Table 4.1 gives the posterior distribution for the population mean \( \mu \) and a Monte Carlo procedure for probabilistic sensitivity analysis when all likelihoods are normal. This posterior distribution is bivariate normal with 2×2 covariance matrix equal to the matrix inverse of the sum \( \sum \Omega_i \), where each \( \Omega_i \) is itself the matrix inverse of the 2×2 matrix determined by the variance estimates \( \sigma^2_{\xi_0}, \sigma^2_{\xi_1}, \sigma^2_{\epsilon_0}, \sigma^2_{\epsilon} \) as indicated in the table. The unknown population mean \( \mu \) has posterior mean vector \( \bar{y} \) given by the product of the covariance matrix \( (\sum \Omega)^{-1} \) with the sum of matrix products \( \Omega y_i \).

The Monte Carlo procedure described in the table requires the generation of a bivariate normal vector \( \mu \). We discuss how to accomplish this in Appendix C.
Table 4.1: Normally distributed observations from heterogeneous controlled studies (random effects model)

<table>
<thead>
<tr>
<th>Parameters for study $i$</th>
<th>Observations $y_i$ for study $i$</th>
<th>Prior distributions</th>
<th>Posterior distribution</th>
<th>Generating new parameter values $\xi^0, \xi^1$ for Monte Carlo simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi^0_i, \xi^1_i$</td>
<td>$y_i^0 \sim \text{normal}(\xi^0_i, \sigma^2_i)$</td>
<td>$\mu = (\mu^0, \mu^1)^T$ has a noninformative prior.</td>
<td>$\mu = (\mu^0, \mu^1)^T \sim \text{normal}(\mathbf{y}, (\sum \Omega)^{-1})$</td>
<td>Step 1. Generate $\mu = (\mu^0, \mu^1)^T \sim \text{normal}(\mathbf{y}, (\sum \Omega)^{-1})$</td>
</tr>
</tbody>
</table>
| $\varsigma^0_i$         | $y_i^1 \sim \text{normal}(\xi^1_i, \sigma^2_i)$ | $\zeta^0|\mu^T \sim \text{normal}(\mu^0, \sigma^2_{\zeta})$. | $\zeta^0 = \left(\begin{array}{cccc}
\sigma^2_{y^0} + \sigma^2_{\xi^0} & -\sigma^2_{y^1}
-\sigma^2_{\xi^0} & \sigma^2_{\xi^1} + \sigma^2_{\xi^1}
\end{array}\right)^{-1}$ | Step 2. Generate $\xi^0 \sim \text{normal}(\mu^0, \sigma^2_{\xi} \omega)$. |
| $\varsigma^1_i$         | $\varsigma^2_i \text{ known or estimated.}$ | $\varsigma^1|\mu^T \sim \text{normal}(\mu^1, \sigma^2_{\varsigma})$. | $\mathbf{y} = (\sum \Omega, y_i)^{-1}(\sum \Omega, y_i)$, | Step 3. Generate $\varsigma \sim \text{normal}(\mu^1, \sigma^2_{\varsigma})$. |
| $\varsigma_{y_i}$       | $\sigma^2_{\varsigma} \text{ known or estimated.}$ | $\sigma^2_{\varsigma_i}$ | $y_j = (y^0_j, y^1_j - y^0_j)^T$ | Step 4. Calculate $\varsigma = \varsigma^0 + \varsigma$. |

Table 4.2 summarizes the corresponding large-sample normal approximations for binomial and Poisson observations in heterogeneous controlled studies.
Table 4.2: Binomial and Poisson observations from heterogeneous controlled studies, and their large-sample normal approximation.

<table>
<thead>
<tr>
<th>Parameters for study $i$</th>
<th>Observations for study $i$</th>
<th>Transformation to an approximate normal random effects model (Table 4.1)</th>
<th>Generating new parameter value: Steps 1,2,3 and 4 are as in Table 4.1. Then:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilities $p^0_i, p^1_i$</td>
<td>$k^0_i$ events in $n^0_i$ independent trials without intervention</td>
<td>$\xi^0_i = \text{logit}(p^0_i)$</td>
<td>Step 5. Calculate $p_0 = \frac{e^{\xi^0_i}}{1 + e^{\xi^0_i}}$, $p_1 = \frac{e^{\xi^1_i}}{1 + e^{\xi^1_i}}$.</td>
</tr>
<tr>
<td></td>
<td>$k^1_i$ events in $n^1_i$ independent trials with intervention</td>
<td>$\xi^1_i = \text{logit}(p^1_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k^0_i \sim \text{binomial}(n^0_i, p^0_i)$</td>
<td>$y^0_i = \text{logit}(k^0_i/n^0_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k^1_i \sim \text{binomial}(n^1_i, p^1_i)$</td>
<td>$y^1_i = \text{logit}(k^1_i/n^1_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2_{i0} = 1/k^0_i + 1/(n^0_i - k^0_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2_{i1} = 1/k^1_i + 1/(n^1_i - k^1_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2 = 1/k_i^0 + 1/(n_i^0 - k^0_i)$</td>
<td></td>
</tr>
<tr>
<td>Rates $\lambda^0_i, \lambda^1_i$</td>
<td>$k^0_i$ events in duration $\Delta t^0_i$ without intervention</td>
<td>$\xi^0_i = \text{log}(\lambda^0_i)$</td>
<td>Step 5. Calculate $\lambda_0 = e^{\xi^0_i}$, $\lambda_1 = e^{\xi^1_i}$, or $RR = e^{\xi}$.</td>
</tr>
<tr>
<td></td>
<td>$k^1_i$ events in duration $\Delta t^1_i$ with intervention</td>
<td>$\xi^1_i = \text{log}(\lambda^1_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k^0_i \sim \text{Poisson}(\lambda^0_i \Delta t^0_i)$</td>
<td>$y^0_i = \text{log}(k^0_i/\Delta t^0_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$k^1_i \sim \text{Poisson}(\lambda^1_i \Delta t^1_i)$</td>
<td>$y^1_i = \text{log}(k^1_i/\Delta t^1_i)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2_{i0} = 1/k^0_i$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\sigma^2_{i1} = 1/k^1_i$</td>
<td></td>
</tr>
</tbody>
</table>

Example 4: The effect of zidovudine prophylaxis on HIV transmission (continuation of Example 2)

In Example 2, we pooled the 7 studies estimating HIV transmission risk with and without zidovudine prophylaxis. In fact, if we calculate the test statistic (1) with the control observations $y^0_i$, we obtain $Q_0 = 39.83$ with p-value $4.9 \times 10^{-7}$. The corresponding statistic $Q_\varepsilon$ using the treatment effects $y^1_i - y^0_i$ gives $Q_\varepsilon = 38.6$ with p-value $8.7 \times 10^{-7}$. The null
Bayesian Posteriors for Probabilistic Sensitivity Analysis

The hypothesis of homogeneous studies is therefore untenable. The corresponding estimates are

\( \hat{\sigma}^2_{\text{obs}} = 0.06911 \) and \( \hat{\sigma}^2_{\epsilon} = 0.2563 \). A random effects analysis using the log transformation from Table 4.2 may be performed as follows.

<table>
<thead>
<tr>
<th>Group ( i )</th>
<th>( y_i^0 = \log(k_i^0 / n_i^0) )</th>
<th>( \sigma_{\text{int}}^2 = 1/k_i^0 - 1/n_i^0 )</th>
<th>( y_i^1 = \log(k_i^1 / n_i^1) )</th>
<th>( \sigma_{\text{int}}^2 = 1/k_i^1 - 1/n_i^1 )</th>
<th>( y_i^1 - y_i^0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.174</td>
<td>0.0147</td>
<td>-3.086</td>
<td>0.05</td>
<td>-1.913</td>
</tr>
<tr>
<td>2</td>
<td>-1.443</td>
<td>0.015</td>
<td>-2.546</td>
<td>0.058</td>
<td>-1.102</td>
</tr>
<tr>
<td>3</td>
<td>-1.153</td>
<td>0.0228</td>
<td>-2.773</td>
<td>0.045</td>
<td>-1.62</td>
</tr>
<tr>
<td>4</td>
<td>-0.9808</td>
<td>0.10417</td>
<td>-1.981</td>
<td>0.216</td>
<td>-0.94</td>
</tr>
<tr>
<td>5</td>
<td>-1.344</td>
<td>0.0246</td>
<td>-1.8</td>
<td>0.044</td>
<td>-0.457</td>
</tr>
<tr>
<td>6</td>
<td>-1.677</td>
<td>0.022</td>
<td>-2.377</td>
<td>0.05</td>
<td>-0.7</td>
</tr>
<tr>
<td>7</td>
<td>-1.6988</td>
<td>0.0008</td>
<td>-2.3199</td>
<td>0.004</td>
<td>-0.621</td>
</tr>
</tbody>
</table>

We obtain

\[
(\sum_i \Omega_i)^{-1} = \begin{pmatrix} 0.013 & -0.0028 \\ -0.0028 & 0.0474 \end{pmatrix}
\]

\[
\sum_i \Omega_i y_i = \begin{pmatrix} -113.2 \\ -28.32 \end{pmatrix}
\]

\[
\bar{y} = \begin{pmatrix} -1.39 \\ -1.02 \end{pmatrix}
\]

We conclude that \( \mu = (\mu^0, \mu^1)^T \) has approximate bivariate normal posterior distribution with mean vector \( \begin{pmatrix} -1.39 \\ -1.02 \end{pmatrix} \) and covariance matrix \( \begin{pmatrix} 0.013 & -0.0028 \\ -0.0028 & 0.0474 \end{pmatrix} \).

We may compare the implied joint posterior distribution on \( p_0, p_1 \) with the implied posterior obtained by (incorrectly) pooling the seven studies as we did in Example 2, and the result is in Figure 5. The posterior distributions are quite different, which is consistent with the strong rejection of the hypothesis of homogeneity that we noted above.

When studies are pooled, the large sample size of study 7 overwhelms the remaining studies, tightens the posterior variance and drags the posterior mean down towards the smaller study-7 mean. Under the heterogeneity assumption, however, study 7 serves as
only one of seven studies, although with a little higher weight due again to its large sample size.

Figure 5. Joint log-normal approximate posterior distributions of HIV vertical transmission risk ($p_0$) without intervention and vertical transmission risk ($p_1$) with zidovudine prophylaxis based on data from seven studies. In (a), the studies are treated as heterogeneous and not pooled. In (b) the studies are (incorrectly) pooled. In addition to different dispersions, note that these posterior distributions have quite different modes, as the scale of graph (b) is only half that of graph (a).

The accuracy of the approximate posterior distribution depends on having large sample sizes. However, the control arm in study 4 appears to fail this requirement, as it contains only $n_i = 16$ observations with $k_i = 3$ transmissions, giving $\hat{p}_i = 0.375$. Nevertheless, as Figure 6 shows, we may be reassured that the normal approximation to the binomial with these parameters is fairly good even though the sample size is small.
Figure 6: Distribution functions of the discrete binomial with $n = 16$, $p = 0.375$, and its normal approximation $\text{Normal}(6,3.75)$.

Combining heterogeneous controlled and uncontrolled studies

It may happen that in addition to controlled studies, there may be additional studies that have no treatment arm. The methods from Table 4.1 and Table 4.2 still apply in this case if they are modified as follows. For studies $i$ without treatment arm, in Table 4.2 set the missing observation $y_i^/$ to an arbitrary value, and omit the calculation of $\sigma_{ii}^2$. Then in Table 4.1 set

$$
\Omega_i = \begin{pmatrix}
\frac{1}{\sigma_{i0}^2 + \sigma_{i0}^2} & 0 \\
0 & 0
\end{pmatrix}
$$

(which is the form taken by $\Omega_i$ when $\sigma_{ii}^2$ is infinitely large). Then apply the procedures in Table 4.1 and Table 4.2 as before. The zero entries in $\Omega_i$ ensure that the arbitrary value of $y_i^/$ does not affect the computation of $\bar{y}$ given in Table 4.1.
An Illustrative Probabilistic Sensitivity Analysis

We now apply these methods to conduct a probabilistic sensitivity analysis for all probability and rate parameters in Mrus and Tsevat’s recently published analysis of zidovudine prophylaxis following rapid HIV testing in labor to prevent vertical HIV transmission in pregnant women\textsuperscript{12}. Figure 7 shows a decision tree depicting this problem, and in Table 5 we list all probability and relative risk parameters used in this model. In the examples above, we have already calculated posterior distributions for four of these parameters. We discuss four remaining parameters here.
Figure 7: The decision to offer rapid testing in labor followed by zidovudine prophylaxis, as modeled by Mrus and Tsevat\textsuperscript{12}. In (a), a mother who accepts rapid HIV testing receives zidovudine upon a positive test, unless delivery occurs before treatment can take effect. The probability of a positive test depends on sensitivity, specificity and HIV prevalence. The subtree (b), in which mother and infant HIV status is revealed, follows each of the terminal nodes in (a). Branch probabilities in (b) depend on the path taken in (a).
Table 5: Probability and relative risk parameters in the Mrus and Tsevat model\textsuperscript{12}. The baseline values listed are values we calculated by transforming posterior means obtained using techniques from Table 1 through Table 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Baseline Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_p$</td>
<td>Prevalence of HIV infection in pregnant women without prenatal care</td>
<td>0.0048</td>
</tr>
<tr>
<td>$P_{t0}$</td>
<td>Vertical transmission risk without intervention (Example 4)</td>
<td>0.2491</td>
</tr>
<tr>
<td>$r_{RR}$</td>
<td>Relative reduction in vertical transmission risk of zidovudine prophylaxis (Example 4)</td>
<td>0.6401</td>
</tr>
<tr>
<td>$P_{ac}$</td>
<td>Test and treatment acceptance rate (Example 1)</td>
<td>0.8558</td>
</tr>
<tr>
<td>$P_d$</td>
<td>Percentage of women delivering before preventative therapy can take effect</td>
<td>0.2746</td>
</tr>
<tr>
<td>$P_{sen}$</td>
<td>Rapid HIV antibody test Sensitivity</td>
<td>0.9962</td>
</tr>
<tr>
<td>$P_{spe}$</td>
<td>Rapid HIV antibody test Specificity (Example 3)</td>
<td>0.9871</td>
</tr>
</tbody>
</table>

Prevalence of HIV infection in pregnant women without prenatal care: Mrus and Tsevat estimate the prevalence of HIV infection in pregnant women in the United States to be $p_0 = 0.0017$ according to a completed survey for the year 1991. They assign a relative risk ($RR$) in the range 2–4 to women without prenatal care. The resulting prevalence of HIV infection without prenatal care is $p = p_0 \cdot RR$.

Because the estimate $p_0$ is based on a very large sample, we take the value $p_0 = 0.0017$ as fixed and include only the relative risk parameter $RR$ in our probabilistic sensitivity analysis. Mrus and Tsevat cite studies by Lindsay et al.\textsuperscript{28} and Donegan et al.\textsuperscript{29} on the relative risk of HIV infection without prenatal care. We obtain a posterior on $RR$ from these studies using Table 4.1 and the log transformation in Table 4.2. The relevant observations are as follows.
Given this data, the approximate posterior distribution of parameter $\mu = (\mu_0, \mu_\varepsilon)^T$ is multivariate normal with mean vector $\left[\begin{array}{c} -4.727 \\ -1.0398 \end{array} \right]$ and covariance matrix $\left[\begin{array}{cc} 0.7966 & -0.0122 \\ -0.0122 & 0.1165 \end{array} \right]$.

We are only interested in the efficacy parameter $\mu_\varepsilon$, the marginal distribution of which is approximate normal with mean $-1.0398$ and variance $0.1165$. The conditional distribution of $\varepsilon = \ln RR$ given $\mu_\varepsilon$ is normal with mean $\mu_\varepsilon$ and variance $\sigma_\varepsilon^2 = 0.1240$, from which it follows that the marginal distribution of $\varepsilon$ is normal with mean $-1.0398$ and standard deviation $(0.1165 + 0.1240)^{1/2} = 0.4904$. The resulting 95% credible interval $-1.0398 \pm 0.9612$ on $\varepsilon$ translates to a 95% credible interval $(1.08, 7.39)$ for $RR$, much wider than the 2–4 range used by Mrus and Tsevat.

**Percentage of women delivering before preventative therapy can take effect:** Mrus and Tsevat estimate the percentage $p$ of women delivering before preventative therapy can take effect as the product $p = p_1 \cdot a$ of the percentage $p_1$ of women without prenatal care who deliver within 4 hours of presentation and the percentage $a$ of women of such women who would deliver before preventative therapy can take effect.

Mrus and Tsevat take $a = \frac{1}{2}$ based on an estimated 2 hours needed to test, treat, and derive benefit from the drug. Our re-analysis uses $a = \frac{1}{2}$ as baseline but for probabilistic sensitivity analysis we take $a$ to have a beta distribution with mean $\frac{1}{2}$ and 95% credible interval 0.25–0.75, that is, a beta(7,7) distribution. Regarding $p_1$, Mrus and Tsevat cite Donegan et al$^{30}$ as observing $k = 306$ of $n = 557$ women without prenatal care delivering.
within 4 hours of presentation. This yields a beta(307,252) posterior on $p_1$, assuming a noninformative prior (see Table 1.1).

Rapid HIV antibody test sensitivity: Mrus and Tsevat cite three studies\textsuperscript{25,26,27} providing data on rapid HIV test sensitivity. These studies, which we pool, observe $k = 262$ positive test results out of $n = 262$ HIV-infected individuals (100% sample sensitivity). Assuming a noninformative uniform prior, this yields a beta(263,1) posterior, (again, see Table 1.1).

Results of probabilistic sensitivity analysis

In our base case analysis, we used the transformed means of the normal posterior distributions for the eight probability and efficacy parameters discussed in this paper, along with Mrus and Tsevat’s estimates for the remaining cost and quality-of-life parameters. This analysis and the subsequent probabilistic sensitivity analysis was conducted in Microsoft Excel using the Stotree software written by one of the authors\textsuperscript{31,32,33,34}. Using the standard value of $50,000/QALY, we found a net benefit of $522.96 per pregnancy for rapid HIV testing followed by zidovudine prophylaxis. This is consistent with the results of Mrus and Tsevat. We performed a probabilistic sensitivity analysis jointly on the eight probability and efficacy parameters discussed in this paper, leaving the remaining cost and quality parameters at the levels specified by Mrus and Tsevat. Our Monte Carlo simulation takes random draws for parameters from their posterior distributions and estimates the probability of decision change and the expected value of perfect information.
Based on 40,000 Monte Carlo iterations, we estimate zidovudine prophylaxis optimal 95.9% (±0.19%) of the time, and the expected value of perfect information on all 8 relative effects and probabilities equal to $10.65 (±$0.87) per pregnancy. These numbers indicate that the optimality of zidovudine prophylaxis is insensitive to simultaneous variation in these eight probability and efficacy parameters, a conclusion consistent with the conventional sensitivity analyses conducted by Mrus and Tsevat. We note, however, that the uncertainty surrounding the cost and quality parameters in this model is substantial, and these should also be included in a probabilistic sensitivity analysis. For the purposes of this paper, however, we confine our analysis to the probability and efficacy parameters.

We also investigated the sensitivity of the optimal decision to variation in each parameter individually, setting all other parameters equal to their transformed normal means. In 10,000 iterations of Monte Carlo simulation for each of the eight parameters, we observed no instances when zidovudine prophylaxis was suboptimal, so both the probability of decision change and the expected value of perfect information are zero for each parameter to within the accuracy of the simulation.

**Conclusion**

We have presented large-sample Bayesian methods for obtaining approximately normal posterior distributions for transformations of probability, rate, and relative-effect parameters given data from either controlled or uncontrolled studies. Means, variances and covariances of these approximating normal distributions can be quickly calculated from study data. Analysts can then generate random variates from these distributions and
reverse transform to obtain probability, rate or relative-effect random variates for the purposes of probabilistic sensitivity analysis. This variate generation and transformation procedure is simple enough to be conducted on a spreadsheet, as we have done. Because these are large-sample approximations, caution must be exercised in applying these techniques to studies with small sample sizes. The results we present may also be used for the calculation of expected value of sample information using Monte Carlo simulation, much in the manner of Ades et al.¹³

Our results assume that population variances within each study – the quantities $\sigma^2$ in Table 1.1, $\sigma_0^2$, $\sigma_1^2$ in Table 2.1, $\sigma_i^2$ in Table 3.1, $\sigma_{i0}^2$, $\sigma_{i1}^2$ Table 4.1 – are well estimated by sample variances, and this is very likely to be so for large samples. We also assume for heterogeneous studies that the variances across studies – the quantity $\sigma_x^2$ in Table 3.1, and the pair of quantities $\sigma_{x0}^2$, $\sigma_{x}^2$ in Table 4.1 – are known or well estimated. However, because the number of studies is typically not large, this is unlikely to be so. A fully Bayesian approach would calculate posterior distributions of these parameters given study data, but such posteriors do not take on a simple form, and may even be improper when a noninformative prior distribution is employed.³⁵ Because we use point estimates for these cross-study variances instead of calculating posterior distributions, the predictive distributions from which we sample during probabilistic sensitivity analysis are too tight to some unknown degree, much in the same way that a normal distribution based on known variance is tighter than the corresponding $t$-distribution that includes uncertainty about variance. The resulting probabilistic sensitivity analysis may therefore underestimate problem sensitivity. A fully Bayesian approach to combining heterogeneous studies would require the specification of a proper prior distribution for
cross-study variance and the use of Markov chain Monte Carlo techniques to estimate posterior distributions.
Appendix A: Proofs for Tables

Lemma\textsuperscript{14,36}: If a random vector $y$ has a multivariate normal distribution $y|\mu \sim \text{Normal}(\mu, \Sigma)$, and $\mu$ is parameterized as $\text{Normal}(\mu_0, \Sigma_0)$, then

i) The posterior on $\mu$ given $y$ is also normal distributed: $\mu|y_1, \ldots, y_n \sim \text{Normal}(\mu_n, \Sigma_n)$.

Where

$$\mu_n = (\Sigma^{-1} + \Sigma_0^{-1})^{-1}(\Sigma^{-1} y + \Sigma_0^{-1} \mu_0), \quad \Sigma_n = (\Sigma^{-1} + \Sigma_0^{-1})^{-1}$$

ii) The marginal distribution of $y$ is $\text{Normal}(\mu_0, \Sigma_0 + \Sigma)$.

Lemma\textsuperscript{15}: (Delta-Method) If $X_n$ is approximately $\text{Normal}(\mu, \sigma_n^2)$, where $\sigma_n$ is a sequence of constants tending to zero, $\mu$ is a fixed number, and $g$ is a differentiable function, then for large $n$, $g(X_n)$ is approximately normal with mean $g(\mu)$ and standard deviation $\sigma_n|g'(\mu)|$.

Theorem 1\textsuperscript{36}: Suppose $y_1, \ldots, y_n$ are $n$ random variables, with $y_i|\zeta_i \sim \text{Normal}(\zeta_i, \sigma_i^2)$, and $y_1, \ldots, y_n$ are independent given $\zeta_1, \ldots, \zeta_n$, where $\zeta_1, \ldots, \zeta_n$ given $\mu$ are i.i.d. random variables, each is $\text{Normal}(\mu, \sigma_{\zeta}^2)$.

i) If $\mu$ has prior distribution $\text{Normal}(\mu_0, \sigma_0^2)$, then the posterior on $\mu$ given $y_1, \ldots, y_n$ is also normal distributed: $\mu|y_1, \ldots, y_n \sim \text{Normal}(\mu_n, \sigma_n^2)$.  

34
Bayesian Posteriors for Probabilistic Sensitivity Analysis

If $\mu$ has noninformative uniform prior distribution, i.e., $f(\mu) \propto 1$, then $\mu$ given $y_1, ..., y_n$ is normal distributed:

$$y_1, ..., y_n \sim \text{Normal}(\bar{y}, \frac{1}{\sum_i \omega_i}).$$

Here $\mu_a = \frac{\sum_i \omega_i \bar{y} + \tau_0 \mu_0}{\sum_i \omega_i + \tau_0}$, $\sigma_a^2 = \frac{1}{\tau_a}$, $\tau_a = \sum_i \omega_i + \tau_0$, and $\omega_i = \frac{1}{\sigma_i^2 + \sigma_0^2}$, $\bar{y} = \frac{\sum_i \omega_i y_i}{\sum_i \omega_i}$,

$$\tau_0 = \frac{1}{\sigma_0^2}.$$

\[\text{Theorem 2: Suppose } y_1, ..., y_n \text{ are } n \text{ (d-dimensional) random vectors, with } y_i | \xi_i \sim \text{Normal}(\xi_i, \Sigma_i), \text{ and } y_1, ..., y_n \text{ are independent given } \xi_1, ..., \xi_n, \text{ where } \xi_1, ..., \xi_n \text{ given } \mu \text{ are i.i.d. (d-dim) random vectors, each is Normal(} \mu, \Sigma_{x0} \text{).} \]

If $\mu$ has prior distribution $\text{Normal}(\mu_0, \Sigma_0)$, then the posterior on $\mu$ given $y_1, ..., y_n$ is also normal distributed: $\mu | y_1, ..., y_n \sim \text{Normal}(\mu_n, \Sigma_n)$.

If $\mu$ has noninformative uniform prior distribution, i.e., $f(\mu) \propto 1$, then $\mu$ given $y_1, ..., y_n$ is normal distributed:

$$y_1, ..., y_n \sim \text{Normal}(\bar{y}, (\sum_i \Omega_i)^{-1}).$$

Here $\mu_a = (\sum_i \Omega_i + \Sigma_x^{-1})^{-1}(\sum_i \Omega_i y_i + \Sigma_x^{-1} \mu_0)$, $\Sigma_a = (\sum_i \Omega_i + \Sigma_x^{-1})^{-1}$, and $\Omega_i = (\Sigma_i + \Sigma_{x0})^{-1}$,

$$\bar{y} = (\sum_i \Omega_i)^{-1}(\sum_i \Omega_i y_i).$$
Pf. Let \( y = (y_1, \ldots, y_n)^T, \xi = (\xi_1, \ldots, \xi_n)^T, \Sigma = \text{diag}\{ \Sigma_1, \ldots, \Sigma_n \}, \Sigma_\xi = \text{diag}\{ \Sigma_{\xi_0}, \ldots, \Sigma_{\xi_0} \} \),

then \( y|\xi \sim \text{Normal}(\xi, \Sigma) \), and \( \xi | \mu \sim \text{Normal}(\mu, \Sigma_\xi) \), \( \mu = (\mu, \ldots, \mu)^T \) is an \( dn \)-dim vector.

So we have \( y|\mu \sim \text{Normal}(\xi, \Sigma + \Sigma_\xi) \), with density

\[
f(y | \mu) \propto \exp \left( -\frac{1}{2} (y - \mu)^T (\Sigma + \Sigma_\xi)^{-1} (y - \mu) \right)
\]

\[
= \exp \left( -\frac{1}{2} \sum_i (y_i - \mu_i)^T (\Sigma_i + \Sigma_{\xi_0})^{-1} (y_i - \mu_i) \right)
\]

\[
= \exp \left( -\frac{1}{2} \sum_i (y_i^T \Omega_i y_i - 2 \mu_i^T \Omega_i y_i + \mu_i^T \Omega_i \mu_i) \right)
\]

\[
\propto \exp \left( -\frac{1}{2} \sum_i (\mu_i^T \Omega_i \mu_i - 2 \mu_i^T \Omega_i y_i) \right)
\]

\[
= \exp \left( -\frac{1}{2} (\mu - \bar{\mu})^T \sum_i \Omega_i (\mu - \bar{\mu}) \right)
\]

If \( \mu \) has prior distribution \( \text{Normal}(\mu_0, \Sigma_0) \), the posterior on \( \mu \) given \( y \) is

\[
f(\mu | y) = f(y | \mu) f(\mu) \propto \exp \left( -\frac{1}{2} (\mu - \bar{\mu})^T \sum_i \Omega_i (\mu - \bar{\mu}) \right) \cdot \exp \left( -\frac{1}{2} (\mu - \mu_0)^T \Sigma_0^{-1} (\mu - \mu_0) \right)
\]

\[
\propto \exp \left( -\frac{1}{2} (\mu - \mu_n)^T \Sigma_n^{-1} (\mu - \mu_n) \right),
\]

where

\[
\mu_n = (\sum_i \Omega_i + \Sigma_0^{-1})(\sum_i \Omega_i \bar{y} + \Sigma_0^{-1})^{-1},
\]

\[
\Sigma_n = (\sum_i \Omega_i + \Sigma_0^{-1})^{-1}.
\]

If \( \mu \) has noninformative uniform prior distribution, then

\[
f(\mu | y) = f(y | \mu) \propto \exp \left( -\frac{1}{2} (\mu - \bar{\mu})^T \sum_i \Omega_i (\mu - \bar{\mu}) \right)
\]
so $\mu \mid y_1, \ldots, y_n \sim \text{Normal}(\bar{y}, (\sum_i \Omega_i)^{-1})$.

Proofs for tables

Since the models in Table 1.1, Table 1.2 and Table 2.1, Table 2.2 can be considered as special and simple case (when $n=1$, and $\mu = \bar{\xi}$) of the models in Table 3.1, Table 3.2, and Table 4.1, Table 4.2, we need only give proofs for the latter Tables.

Proofs for Table 3.1, Table 3.2

Table 3.1 is immediately from Theorem 1, since $y_1, \ldots, y_n$ are $n$ random variables, with $y_i \mid \xi_i \sim \text{Normal}(\xi_i, \sigma_i^2)$, and $y_1, \ldots, y_n$ are independent given $\xi_1, \ldots, \xi_n$, where $\xi_1, \ldots, \xi_n$ given $\mu$ are i.i.d. random variables, each is $\text{Normal}(\mu, \sigma_\xi^2)$. So, if $\mu$ has noninformative uniform prior distribution, then $\mu \mid y_1, \ldots, y_n \sim \text{Normal}(\bar{y}, \frac{1}{\sum_i \omega_i})$.

In Row 1 of Table 3.2, we have $n$ study groups, and each study $i$, $i=1,2\ldots n$, has $k_i$ successes in $n_i$ independent trials: $k_i \sim \text{binomial}(n_i, p_i)$, where $p_i$ is the ‘true’ probability estimated by study $i$. And $y_i = \text{logit}(k_i/n_i)$, $\xi_i = \text{logit}(p_i)$.

We want to show that each observed $y_i$ has approximately normal distribution about their expected value $\xi_i$, i.e., $y_i \mid \xi_i$ is approximately $\text{Normal}(\xi_i, \sigma_i^2)$, moreover, $\sigma_i^2$ is estimated by $\sigma_i^2 = 1/k_i + 1/(n_i-k_i)$.

Since $k_i \mid p_i \sim \text{binomial}(n_i, p_i)$, by the normal approximation of binomial distribution (when $n_i$ is large), $(k_i/n_i) \mid p_i$ approximately $\text{Normal}(p_i, \frac{1}{n_i} p_i (1 - p_i))$. For $y_i = \text{logit}(k_i/n_i)$, by the Delta-Method above, straightforward computations show that $y_i \mid \xi_i$ is
Bayesian Posteriors for Probabilistic Sensitivity Analysis

approximately Normal(\(\zeta_i, \sigma_i^2 = \frac{1}{n_i} p_i (1 - p_i) \cdot (\text{logit}(p_i))^2 = \frac{1}{n_i p_i (1 - p_i)}\)). By replacing \(p_i\) by its estimate \(k_i/n_i\), we use \(1/k_i + 1/(n_i-k_i)\) as an estimator of \(\sigma_i^2\).

Row 2 of Table 3.2 is a similar model as in Row 1, and the only difference is \(y_i = \log(k_i/n_i)\), \(\zeta_i = \log(p_i)\) (instead of \(y_i = \text{logit}(k_i/n_i), \zeta_i = \text{logit}(p_i)\)). We get by Delta-Method that \(\sigma_i^2 = 1/k_i - 1/n_i\) (instead of \(\sigma_i^2 = 1/k_i + 1/(n_i-k_i)\)).

In Row 3 of Table 3.2, each study \(i, i=1,2\ldots n\), has \(k_i\) occurrences in duration \(\Delta t_i\), where \(k_i \sim \text{Poisson}(\lambda_i \Delta t_i)\). And \(y_i = \log(k_i/\Delta t_i), \zeta_i = \log(\lambda_i)\). We want to show that \(y_i | \zeta_i\) is approximately Normal(\(\zeta_i, \sigma_i^2\)). Here, \(\sigma_i^2 = 1/k_i\).

Since \(k_i | \lambda_i \sim \text{Poisson}(\lambda_i \Delta t_i)\). by the normal approximation of Poisson distribution (when \(\Delta t_i\) is large), \((k_i/\Delta t_i) | \lambda_i\) approximately Normal(\(\lambda_i, \lambda_i/\Delta t_i\)). For \(y_i = \log(k_i/\Delta t_i)\), using delta-Method (the Lemma above), \(y_i | \zeta_i\) is approximately Normal(\(\zeta_i, \sigma_i^2 = 1/(\lambda_i \Delta t_i)\)). Since \(k_i/\Delta t_i\) is an estimator of \(\lambda_i\), we use \(1/k_i\) as an estimator of \(\sigma_i^2\).

Proofs for Table 4.1, Table 4.2

Table 4.1 is immediate from Theorem 2: Since \(y_1,\ldots, y_n\) are \(n\) (2-dim)random vectors, with

\[ y_i \mid \xi_i \sim \text{Normal}(\xi_i, \Sigma_i), \text{ where } \Sigma_i = \begin{pmatrix} \sigma_{i0}^2 & -\sigma_{i0}^2 \\ -\sigma_{i0}^2 & \sigma_{ii}^2 \end{pmatrix}, \text{ and } y_1,\ldots, y_n \text{ are independent given } \xi_1,\ldots, \xi_n, \text{ where } \xi_1,\ldots, \xi_n \text{ are i.i.d. (2-dim) random vectors, each is } \text{Normal}(\mu, \Sigma_{\xi_0}), \text{ with } \Sigma_{\xi_0} = \begin{pmatrix} \sigma_{\xi0}^2 & 0 \\ 0 & \sigma_{\xi}^2 \end{pmatrix}. \text{ So if } \mu \text{ has noninformative uniform prior }
Bayesian Posteriors for Probabilistic Sensitivity Analysis

distribution, then \( \mu | y_1, ..., y_n \sim \text{Normal}(\bar{y}, (\sum_i \Omega_i)^{-1}) \), where \( \Omega_i = (\Sigma_i + \Sigma_{x0})^{-1} = \left( \begin{array}{cc} \sigma^2_{i0} + \sigma^2_{x0} & -\sigma^2_{i0} \\ -\sigma^2_{i0} & \sigma^2_{i0} + \sigma^2_{x} \end{array} \right)^{-1} \).

Row 1 of Table 4.2, we have \( n_i \) study groups, and each study \( i, i=1,2..., n \), has \( k_i^0 \) successes in \( n_i^0 \) independent trials without intervention, and \( k_i^l \) successes in \( n_i^l \) independent trials with intervention. \( k_i^0 \sim \text{binomial}(n_i^0, p_i^0), k_i^l \sim \text{binomial}(n_i^l, p_i^l), k_i^0, k_i^l \) are independent given \( p_i^0, p_i^l \). And \( y_i^0 = \logit(k_i^0/n_i^0), \xi_i^0 = \logit(p_i^0), y_i^l = \logit(k_i^l/n_i^l), \xi_i^l = \logit(p_i^l), \eta_i = y_i^l - y_i^0, \epsilon_i = \xi_i^l - \xi_i^0 \).

We want to show that each observed \( y_i^0, y_i^l \) have approximately normal distributed about their expected value \( \xi_i, \epsilon_i \). Let \( y_i = (y_i^0, x_i)^T, \xi_i = (\xi_i^0, \xi_i^l)^T \) then we want to show that \( y_i | \xi_i \) is approximately Normal(\( \xi_i, \Sigma_i \)), moreover, \( \Sigma_i \) is estimated by \( \Sigma_i = \left( \begin{array}{cc} \sigma^2_{i0} & -\sigma^2_{i0} \\ -\sigma^2_{i0} & \sigma^2_{i0} + \sigma^2_{i} \end{array} \right) \), \( \sigma^2_{i0} = 1/k_i^0 + 1/(n_i^0 - k_i^0) \), \( \sigma^2_{i} = 1/k_i^l + 1/(n_i^l - k_i^l) \).

By Row 1 of Table 3.2, we have \( y_i^0 | \xi_i^0 \) is approximately Normal(\( \xi_i^0, \sigma^2_{i0} \)), and \( y_i^l | \xi_i^l \) is approximately Normal(\( \xi_i^l, \sigma^2_{i} \)), where \( \sigma^2_{i0}, \sigma^2_{i} \) are estimated by \( \sigma^2_{i0} = 1/k_i^0 + 1/(n_i^0 - k_i^0) \), \( \sigma^2_{i} = 1/k_i^l + 1/(n_i^l - k_i^l) \).

We have assumed that \( y_i^0, y_i^l \) are independent given \( \xi_i^0, \xi_i^l \). Note that \( x_i = y_i^l - y_i^0 \) and \( y_i^0 \) are negative linearly dependent, \( \text{Cov}(x_i, y_i^0) = -\sigma^2_{i0} \). So we get \( \Sigma_i = \left( \begin{array}{cc} \sigma^2_{i0} & -\sigma^2_{i0} \\ -\sigma^2_{i0} & \sigma^2_{i0} + \sigma^2_{i} \end{array} \right) \).

Row 2 of Table 4.2 is a similar model as in Row 1, and the only difference is \( y_i^0 = \log(k_i^0/n_i^0), \xi_i^0 = \log(p_i^0), y_i^l = \log(k_i^l/n_i^l), \xi_i^l = \log(p_i^l) \).
logit(\(p_i^0\)), \(y_i^0 = logit(k_i^0 / n_i^0\)), \(\xi_i^0 = logit(p_i^0)\). Using the result in Row 2 of Table 3.2, and by the same argument as above, we get
\[
\Sigma_i = \begin{pmatrix}
\sigma_{i0}^2 & -\sigma_{i0}^2 \\
-\sigma_{i0}^2 & \sigma_{i0}^2 + \sigma_{i1}^2
\end{pmatrix}
\]
and \(\sigma_{i0}^2 = 1/k_i^0 - 1/n_i^0\).

In Row 3 of Table 4.2, each study \(i, i=1,2,\ldots n\), has \(k_i^0\) occurrences in duration \(\Delta t_i^0\) without intervention, and \(k_i^1\) occurrences in duration \(\Delta t_i^1\) with intervention. \(k_i^0 \sim \text{Poisson}(\lambda_i^0 \Delta t_i^0)\), \(k_i^1 \sim \text{Poisson}(\lambda_i^1 \Delta t_i^1)\), \(k_i^0, k_i^1\) are independent given \(\lambda_i^0, \lambda_i^1\). And \(y_i^0 = \log(k_i^0/\Delta t_i^0), \xi_i^0 = \log(\lambda_i^0), y_i^1 = \log(k_i^1/\Delta t_i^1), \xi_i^1 = \log(\lambda_i^1), x_i = y_i^1 - y_i^0, \epsilon_i = \xi_i^1 - \xi_i^0\).

Also we want to show that each observed \(y_i^0, x_i\) are approximately normal distributed about their expected value \(\xi_i^0, \epsilon_i\). Let \(y_i=(y_i^0, x_i)^T, \xi_i=(\xi_i^0, \epsilon_i)^T\), then we want to show that \(y_i|\xi_i\) is approximately Normal(\(\xi_i, \Sigma_i\)), moreover, \(\Sigma_i\) is estimated by
\[
\Sigma_i = \begin{pmatrix}
\sigma_{i0}^2 & -\sigma_{i0}^2 \\
-\sigma_{i0}^2 & \sigma_{i0}^2 + \sigma_{i1}^2
\end{pmatrix}
\]
\(\sigma_{i0}^2 = 1/k_i^0, \sigma_{i1}^2 = 1/k_i^1\). The argument is also the same as above, by using the result in Row 3 of Table 3.2.

**Appendix B: Models with Unbalanced Data**

If there are \(n\) control groups and \(m\) treatment groups, with \(n>m\), using the same notations as before, the model could be explained as follows: There are \(n-m\) study groups (which do not contain a treatment arm) giving data \(y_i^0\) which estimate parameters \(\xi_i^0, i=1,2,\ldots n-m\).

For these studies, \(y_i^0 \sim \text{Normal}(\xi_i^0, \sigma_{i0}^2)\), and variance \(\sigma_{i0}^2\). And \(\xi_i^0, \ldots, \xi_{n-m}^0\) given \(\mu^0\) are i.i.d. random variables, each is Normal(\(\mu^0, \sigma_{i0}^2\)). In the other \(m\) study groups with treatment arms, we have \(y_i=(y_i^0, x_i)^T, \xi_i=(\xi_i^0, \epsilon_i)^T\), where \(x_i = y_i^1 - y_i^0, \epsilon_i = \xi_i^1 - \xi_i^0\). Here \(y_i|\xi_i\) is approximately Normal(\(\xi_i, \Sigma_i\))...
where $\Sigma_i$ is estimated by $\Sigma_i = \begin{bmatrix} \sigma_{i0}^2 & -\sigma_{i0}^2 \\ -\sigma_{i0}^2 & \sigma_{i0}^2 + \sigma_{ii}^2 \end{bmatrix}$. And $\xi_{n-m+1, \ldots, n}$ given $\mu = (\mu^0, \mu^\varepsilon)^T$ are i.i.d. (2-dim) random vectors, each is Normal($\mu, \Sigma_{\xi_0}$), with $\Sigma_{\xi_0} = \begin{bmatrix} \sigma_{\xi_0}^2 & 0 \\ 0 & \sigma_{\varepsilon}^2 \end{bmatrix}$.

A Bayesian approach to obtain the posterior for $\mu = (\mu^0, \mu^\varepsilon)^T$ would be as follows: Under the assumption of noninformative uniform prior distribution for $\mu = (\mu^0, \mu^\varepsilon)^T$, update the distribution for $\mu^0$ (using the methods developed in Table 3.1 and Table 3.2) according to the data from the $(n-m)$ groups which do not contain studies for treatment. Then use this distribution for $\mu^0$ and noninformative distribution for $\mu^\varepsilon$ as the prior to obtain the posterior distribution for $\mu = (\mu^0, \mu^\varepsilon)^T$ given the data from the remaining $m$ control groups and $m$ treatment groups (using the conclusion in Theorem 2).

This approach leads to the same result as the simpler method we presented in the main part of this paper, which will be shown in the following Theorem.

**Theorem 3**: For a random effects model with unbalanced data, which is discussed above, suppose $\mu = (\mu^0, \mu^\varepsilon)^T$ has noninformative uniform prior distribution, then $\mu \mid y_j, \ldots, y_n \sim$ Normal($\bar{y}, (\sum_{i=1}^{n} \Omega_i)^{-1}$), where

$$\bar{y} = (\sum_{i=1}^{n} \Omega_i)^{-1} (\sum_{i=1}^{n} \Omega_i y_i), \quad \Omega_i = \begin{bmatrix} \omega_i & 0 \\ 0 & 0 \end{bmatrix}, \quad \omega_i = \frac{1}{\sigma_{i0}^2 + \sigma_{\xi_0}^2}, \quad i=1,2,\ldots,n-m,$$

and

$$\Omega_i = (\Sigma_i + \Sigma_{\xi_0})^{-1} = \begin{bmatrix} \sigma_{i0}^2 + \sigma_{\xi_0}^2 & -\sigma_{i0}^2 \\ -\sigma_{i0}^2 & \sigma_{i0}^2 + \sigma_{ii}^2 + \sigma_{\varepsilon}^2 \end{bmatrix}^{-1}, \quad i=n-m+1,\ldots,n.$$
First using Theorem 1, we get the posterior of $\mu_0$ given data $y_i^0$, $i=1,2,...,n-m$, which is Normal($\tilde{\mu}_0$, $\Sigma_0$), where $\tilde{\mu}_0 = \frac{1}{\sigma_0^2} \sum_{i=1}^{n-m} \omega_i y_i^0$, $\sigma_0^2 = \frac{1}{\sigma_i^2 + \sigma_0^2}$.

Then we use this to update the prior of parameter $\mu = (\mu_0, \mu^\epsilon)^T$. Since $\mu^\epsilon$ has a noninformative distribution, or Normal(0, $\sigma^2$) with $\sigma^2 \approx \infty$, the distribution of $\mu = (\mu_0, \mu^\epsilon)^T$ is multivariate normal($\mu_0$, $\Sigma_0$), where $\mu_0 = \begin{pmatrix} -0 \\ 0 \end{pmatrix}$, $\Sigma_0 = \begin{pmatrix} \omega_i \\ 0 \\ 0 \end{pmatrix}$, and $\Sigma_0^{-1} = \begin{pmatrix} \sum_{i=1}^{n-m} \omega_i & 0 \\ 0 & \sigma^2 \end{pmatrix}$. Hence $\Sigma_0^{-1} = \sum_{i=1}^{n-m} \Omega_i$.

Applying Theorem 2, the posterior on $\mu$ given $y_{n-m+1},...,y_n$ is normally distributed: $\mu | y_1,...,y_n \sim \text{Normal}(\mu_n, \Sigma_n)$, where

$$\Sigma_n = (\sum_{i=n-m+1}^{n} \Omega_i + \Sigma_0^{-1})^{-1} = (\sum_{i=n-m+1}^{n} \Omega_i) + \sum_{i=1}^{n-m} \Omega_i = (\sum_{i=1}^{n} \Omega_i)^{-1}$$

and

$$\mu_n = (\sum_{i=n-m+1}^{n} \Omega_i + \Sigma_0^{-1})^{-1} (\sum_{i=n-m+1}^{n} \Omega_i y_i + \Sigma_0^{-1} \mu_0)$$

$$= (\sum_{i=1}^{n} \Omega_i)^{-1} (\sum_{i=n-m+1}^{n} \Omega_i y_i + \sum_{i=1}^{n-m} \omega_i y_i^0)$$

This completes the proof.
Appendix C: Bivariate Normal Random Variate Generation

**Proposition**: (From joint to conditional) If the joint distribution of random vector \( \begin{pmatrix} x \\ y \end{pmatrix} \) is multivariate normal with mean \( \begin{pmatrix} m_x \\ m_y \end{pmatrix} \) and covariance matrix \( \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \), then the conditional distribution of \( y \) given \( x \) is multivariate normal with conditional mean \( \mathbb{E}[y|\mathbf{x}] = m_{y|x} = m_y + \Sigma_{yx} \Sigma_{xx}^{-1}(x - m_x) \) and conditional covariance matrix \( \Sigma_{yy|x} = \Sigma_{yy} - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \).

Therefore, when \( \mu = (\mu^0, \mu^\varepsilon)^T \) is bivariate normal with mean \( (m_1, m_2)^T \) and covariance matrix \( \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \), then we can generate a bivariate normal vector \( \mu \) by 2 steps:

1. **Step 1**: Generate \( \mu^0 \) from \( \mathcal{N}(m_1, \sigma_1^2) \), i.e., normal distribution with mean \( m_1 \) and variance \( \sigma_1^2 \).

2. **Step 2**: Generate \( \mu^\varepsilon \) from \( \mathcal{N}(m_2 + (\mu^0 - m_1)\sigma_{21} / \sigma_1^2, \sigma_2^2 - \sigma_{12} \sigma_{21} / \sigma_1^2) \), which is the conditional distribution of \( \mu^\varepsilon \) when \( \mu^0 \) is given.
References


33 Hazen GB. Stochastic trees and the StoTree modeling environment: models and software for medical decision analysis”, Journal of Medical Systems 2002; 26:399-413.

34 Hazen GB. Stochastic trees and the StoTree modeling software.

http://users.iems.nwu.edu/~hazen/

