Practice of Epidemiology

Bayesian Posterior Distributions Without Markov Chains

Stephen R. Cole*, Haitao Chu, Sander Greenland, Ghassan Hamra, and David B. Richardson

* Correspondence to Dr. Stephen R. Cole, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Campus Box 7435, Chapel Hill, NC 27599-7435 (e-mail: cole@unc.edu).

Bayesian posterior parameter distributions are often simulated using Markov chain Monte Carlo (MCMC) methods. However, MCMC methods are not always necessary and do not help the uninitiated understand Bayesian inference. As a bridge to understanding Bayesian inference, the authors illustrate a transparent rejection sampling method. In example 1, they illustrate rejection sampling using 36 cases and 198 controls from a case-control study (1976–1983) assessing the relation between residential exposure to magnetic fields and the development of childhood cancer. Results from rejection sampling (odds ratio (OR) = 1.69, 95% posterior interval (PI): 0.57, 5.00) were similar to MCMC results (OR = 1.69, 95% PI: 0.58, 4.95) and approximations from data-augmentation priors (OR = 1.74, 95% PI: 0.60, 5.06). In example 2, the authors apply rejection sampling to a cohort study of 315 human immuno-deficiency virus seroconverters (1984–1998) to assess the relation between viral load after infection and 5-year incidence of acquired immunodeficiency syndrome, adjusting for (continuous) age at seroconversion and race. In this more complex example, rejection sampling required a notably longer run time than MCMC sampling but remained feasible and again yielded similar results. The transparency of the proposed approach comes at a price of being less broadly applicable than MCMC.

Bayes theorem; epidemiologic methods; inference; Monte Carlo method; posterior distribution; simulation

Bayesian posterior parameter distributions are usually simulated using Markov chain Monte Carlo (MCMC) methods (1–6). In some simple settings, one may directly calculate the posterior parameter distribution of interest without the need for MCMC methods. However, when a model contains many parameters (as when it contains many confounders), direct calculations can become intractable, and approximations become necessary. Those approximations can be divided into 2 types: posterior simulation, in which a picture of the posterior distribution is built by sampling from it, and analytic approximation, in which properties of the posterior distribution (such as its mode, mean, and variance) are described using large-sample formulas similar to those used in frequentist inference to describe likelihood summaries (such as the maximum-likelihood estimator).

Analytic approximations can be done quickly in comparison with simulations, and thus they facilitate large-scale sensitivity analyses. As another advantage, analytic approximations based on data augmentation can be conducted using ordinary regression software (7–9). This is because data augmentation represents the prior by means of pseudodata, which can be combined with the observed study data using methods familiar to epidemiologists, such as inverse-variance weighting or maximum likelihood. In this way, data augmentation provides an introduction to Bayesian inference that is quite natural to epidemiologists and easy to carry out with whatever software is at hand.

Posterior simulation methods have important advantages, however. Even in small studies, the simulation error can be

made arbitrarily small by making the number of simulation draws large enough (albeit at the cost of longer run time); in this sense, simulation provides a Bayesian analog to frequentist exact methods. Furthermore, unlike basic analytic approximations, MCMC simulation can handle very high-dimensional problems that often arise today (e.g., in genomics). It is thus unsurprising that, over the past few decades, MCMC methods have been the most popular simulation techniques among statisticians.

Unfortunately, due largely to the fact that MCMC produces correlated draws from the posterior distribution, the procedure can run very slowly, requires more technical attention to convergence issues than do other methods, and requires more sophisticated theory to understand its inner workings. To provide a bridge for understanding sophisticated methods such as MCMC, we illustrate a simple rejection-sampling approach (10) to simulate posterior distributions and compare it with other methods using 2 real-data examples. We conclude that, despite limitations, rejection sampling is useful for teaching and implementation of simple Bayesian analysis in epidemiology.

MATERIALS AND METHODS

Example 1: magnetic fields and childhood cancer

In a classic and somewhat controversial early study of the relation between residential exposure to magnetic fields and the development of childhood cancer, Savitz et al. (11) collected data on all 356 cancers diagnosed in residents under age 15 years between 1976 and 1983 in the 5-county 1970 Denver, Colorado, Standard Metropolitan Statistical Area. Controls were selected by random digit dialing. Exposure was assessed using in-home electric and magnetic field measurements under low and high power conditions. To facilitate direct comparison between the approach illustrated here and data augmentation as described by Greenland (8), we restrict our attention to the 36 leukemia cases and 198 controls examined previously (12). We concentrate on magnetic field exposures classified as greater than 3 mG (milligauss) under low power-use conditions ($x = 1$ for $\geq$3 mG, 0 for <3 mG), leading to 3 exposed cases and 5 exposed controls. The data model we will use assumes that the probability π_i of being a leukemia case is $\pi_i = \text{expit}(\beta_0 + \beta_1 x_i)$, where expit(.) = $\exp(.)/[1 + \exp(.)]$ and $\exp(\beta_1)$ is the odds ratio for the association of fields greater than 3 mG with leukemia.

Example 2: viral load and incident AIDS

The Multicenter AIDS Cohort Study (13) began in 1984 and enrolled 6,972 homosexual and bisexual men in Baltimore, Maryland; Chicago, Illinois; Pittsburgh, Pennsylvania; and Los Angeles, California. Every 6 months, participants underwent a physical examination, completed an extensive interviewer-administered questionnaire collecting information on use of antiretroviral therapy, and provided a blood sample for the determination of CD4 cell count and human immunodeficiency virus (HIV) viral load. Positive enzyme-linked immunosorbent assays with confirmatory Western blots were used to determine seropositivity for HIV type 1.

Data were obtained from version 11 of the public use data set and were restricted to 315 men who were observed to seroconvert (become HIV-positive) between 1984 and 1998. Sixty-one incident cases of acquired immunodeficiency syndrome (AIDS) were observed within 5 years of HIV seroconversion; 19 men were lost to follow-up, and 235 men were administratively censored at the minimum of 5 years or 1998. The exposure x was a high viral load set point, defined as a viral load greater than 10^5 copies/mL of plasma measured within 9 months of the estimated seroconversion date. Covariates included race (50 nonwhites, 265 whites; $r = 1$ indicates nonwhite race) and age at seroconversion (median, 34 years; quartiles, 29 and 39 years; a_i = age in decades). Of 315 men, 73 were exposed with a high viral load set point, and 25 of these 73 exposed men were AIDS cases. The data model we will use assumes that the probability π_i of being an AIDS case for patient i is $\pi_i = \text{expit}(\beta_0 + \beta_1 x_i + \beta_2 a_i + \beta_3 r_i)$, where $\exp(\beta_1)$ is the odds ratio for the association of high viral load with AIDS.

Prior specification

Bayes' theorem states that the posterior distribution for parameters of interest β given observed data O, $f(\beta|O)$, is proportional to the prior distribution for β, $f(\beta)$, multiplied by the likelihood of the observed data $L(\beta; O)$: $f(\beta|O) \propto f(\beta) \times L(\beta; O)$. In example 1, the parameters $\beta = (\beta_0, \beta_1)$ include the intercept β_0 and the log odds ratio β_1, and the data $O = \{X,Y\}_i$ for $i = 1$ to 234 consist of an indicator X of exposure to greater than 3 mG and an indicator Y of leukemia; the likelihood is then of the binomial form $L(\beta; O) = \prod_{i=1}^{n} \pi_i^{y_i} \times (1 - \pi_i)^{(1-y_i)}$.

In example 2, the data $O = \{A,R,X,Y\}_i$ for $i = 1$ to 315 consist of age and indicators for nonwhite race ($r = 1$), exposure to a high viral load set point ($x = 1$), and incident AIDS ($y = 1$), respectively. The likelihood also follows the binomial form given above, where $\pi_i = \text{expit}(\beta_0 + \beta_1 x_i + \beta_2 a_i + \beta_3 r_i)$.

In example 1, following Greenland (8), we use a null-centered lognormal prior for the magnetic field-leukemia odds ratio with 95% of the prior mass between an odds ratio of 1/4 and 4, representing nondirectional prior information that the association is probably small or at most modest. This prior corresponds to a normal distribution for the log odds ratio β_1 with a mean (location parameter) $\mu = 0$ and a variance $\sigma^2 = 1/2$ (so that σ is the prior standard deviation of β_1), and is roughly equivalent to adding a stratum with 4 exposed cases and 4 unexposed cases to the data (8). We use a vague prior for the intercept—specifically, a normal distribution for β_0 with a mean $\mu = 0$ and a variance $\sigma^2 = 100$ (use of an even less informative prior on the intercept (i.e., $\sigma^2 = 1,000$) did not alter inferences). We assume that the priors for the log odds ratio and intercept are independent.

In example 2, we again assume the same vague prior for the intercept β_0 and the same null-centered moderately informative log odds ratio prior for β_1, the coefficient of the high-viral-load set point exposure (i.e., β_1 has prior mean $\mu = 0$ and prior variance $\sigma^2 = 1/2$). We also assign the same normal (0, 1/2) prior to the race coefficient β_3 (14). However, given existing evidence of higher AIDS incidence with older age (15, 16), the normal prior for the age coefficient β_2 was chosen to give an odds ratio of 1.25 per decade of age with

95% of the prior mass between 0.84 and 1.85 (implied by giving β_2 prior mean $\mu = 0.223$ and prior variance $\sigma^2 = 1/25$). Again, we assume independent priors for all of the coefficients.

Maximum-likelihood and MCMC methods

All analyses were conducted using SAS, version 9.2 (SAS Institute, Inc., Cary, North Carolina). Maximum-likelihood results were obtained from the SAS procedure GENMOD, which computes confidence intervals using the Wald method (i.e., estimate $\pm$ 1.96 times the approximate standard error). For the initial Bayesian analysis, we generated $20{,}000 \times K$ MCMC draws from the posterior distribution using the BAYES statement in the SAS procedure GENMOD, with the priors described above. We discarded the initial 1,000 draws for each chain to help ensure that the sampler had converged before obtaining the $20{,}000 \times K$ draws that we used. Details about the MCMC approach used by SAS are provided in Appendix 1. In both examples, we chose $K = 5$ and used every fifth draw to minimize autocorrelation of the draws; this choice produced absolute autocorrelations below 0.01. We also computed the Gelman-Rubin convergence diagnostic using 3 chains, each of size $20{,}000 \times K$ (17); in both examples, this diagnostic was essentially 1, which suggests that nonconvergence was not detected. In Appendix 1, we also provide SAS code for implementing this MCMC approach.

Rejection sampling

The rejection-sampling approach we illustrate has 4 steps: 1) draw parameters from the joint prior distribution of the model parameters; 2) perform a standard maximum-likelihood analysis of the observed data; 3) compute the likelihood of each prior draw given the observed data, relative to the maximum (i.e., the acceptance ratio, also sometimes called the importance ratio); and 4) accept a draw with a probability based on the relative likelihood of the draw. We provide a step-by-step overview of the process using example 1.

We begin by generating a sample of M draws from the joint prior distribution of the two parameters, namely the intercept β_0 and log odds ratio β_1. The range of the prior must contain the range of the posterior or the rejection sampler will not fully capture the posterior distribution. To protect against extremely long run times, we set a maximum M of 10,000,000 (which is necessary in example 2).

Second, we find the maximum-likelihood estimates from the observed data and compute the maximum of the likelihood $L(\hat{\beta}; O)$, where $\beta = (\beta_0, \beta_1)$ is the vector of all of the model parameters; this maximum is just the antilog of the model log likelihood supplied by the logistic regression program.

Third, we compute the likelihood for each of the M prior draws, $L(\beta_m; O)$, using the likelihood function from the observed data. Specifically, these are the values of the likelihood function at the prior parameter values that were drawn. From these likelihood values, we compute the relative likelihood or acceptance ratio, $p_m = L(\beta_m; O)/L(\hat{\beta}; O)$, which has the range 0–1.

Fourth and lastly, for each draw $m = 1$ to M, we compare a uniform random number U_m, drawn from the range 0–1, against the acceptance ratio p_m. Specifically, we select only those parameter draws for which $U_m < p_m$ as draws from the posterior distribution. Therefore, the lower the likelihood of a prior parameter draw, the more probable is its rejection from the posterior sample. SAS code for implementing this rejection sampling for example 1 is given in Appendix 2.

Data augmentation

We implemented data augmentation using offsets, with rescaling to improve the asymptotic approximation. Because details are covered elsewhere by Greenland (9), we provide only a brief outline.

One added data record was constructed for each prior, including the intercept. These prior records were appended to the observed data. Additionally, an offset term was added to the regression model, and an explicit intercept was added to the covariate vector (with the automatic intercept suppressed). The offset term is 0 for actual data records. Each prior record represents a subgroup with A pseudocases out of $2A$ pseudo-observations. The offset for this record is given by $-\ln(\mu)/S$, where $S = \sigma(A/2)^{1/2}$ is a rescaling factor. We chose $A = 500$, which leads to $S \approx 11.18$ when $\sigma^2 = 1/2$ as for the target log odds ratio β_1. The exposure level in this record is set to $X = 1/S \approx 1/11.18 = 0.0894$, while all other covariates (including the intercept) are set to 0. In Appendix 1, we provide SAS code for implementing this form of data augmentation.

Bayesian statistical summaries

For rejection sampling and MCMC approaches, we present the mean of the posterior draws and the approximate 95% posterior interval computed from the mean $\pm$ 1.96 times the standard deviation of the draws. We refer to these as the Wald limits, and they assume posterior normality. We also provide the median values and 2.5th and 97.5th percentiles of the draws, which do not assume normality but are more sensitive to simulation variability. We estimate the simulation standard deviation by splitting the posterior draws into 5 equal-sized blocks, calculating the standard deviation of the 5 block-specific means of β_1, and dividing by $5^{1/2}$.

For data augmentation, we present the approximate posterior mean and 95% posterior interval computed by analyzing the augmented data (actual observations and pseudo-observations) using maximum-likelihood logistic regression (9). Thus, the approximate mean is now the posterior mode (i.e., the point at which the posterior is maximized), and the Wald posterior limits are this point $\pm$ 1.96 times the estimated standard deviation computed from the inverse of the total (i.e., prior and likelihood) information matrix. This approach also assumes posterior normality; this assumption can be weakened by computing penalized-deviance limits, which are profile-likelihood limits from the full augmented data set (18).

RESULTS

Example 1: magnetic fields and childhood cancer

Table 1 presents results from several analyses of the leukemia case-control data (8). First shown is the maximum-likelihood estimate of the log odds ratio and its standard

Table 1.   Estimated Odds Ratios for Childhood Leukemia According to Residential Exposure to Magnetic Fields Above 3 mG, Denver, Colorado, 1976–1983[a]

Method	Estimate of β_1	SE or SD[b]	OR = $\exp(\beta_1)$	95% CI or 95% PI[c]
Maximum likelihood	1.255	0.754[d]	3.51	0.80, 15.4[e]
Bayesian[f]				
MCMC[g]	0.527	0.546	1.69	0.58, 4.95
	0.537		1.71	0.57, 4.76
Rejection sampling[g]	0.526	0.553	1.69	0.57, 5.00
	0.534		1.71	0.56, 4.90
Data augmentation[h]	0.555	0.544	1.74	0.60, 5.06

Abbreviations: CI, confidence interval; MCMC, Markov chain Monte Carlo; OR, odds ratio; PI, posterior interval; SD, standard deviation; SE, standard error.

[a] Data were obtained from a case-control study by Savitz et al. (11).

[b] Standard deviation unless otherwise specified.

[c] 95% posterior interval unless otherwise specified.

[d] Standard error.

[e] 95% confidence interval.

[f] All Bayesian methods used a lognormal odds-ratio prior with 95% limits of 1/4 and 4.

[g] First estimate of β_1 is the mean of 20,000 draws; first limits are $\exp(\text{mean} \pm 1.96 \times \text{SD})$, where the mean and SD are computed from the draws. Second estimate is the median; second limits are the 2.5th and 97.5th percentiles. The simulation error was 0.0053 for MCMC and 0.0034 for rejection sampling.

[h] Estimate is the posterior mode (maximum of posterior); limits are $\exp(\text{mean} \pm 1.96 \times \text{SD})$, where the SD is the β_1 diagonal entry from the inverse of the total (i.e., data and prior) information matrix.

error. These yield an estimate of the odds ratio $\exp(\beta_1)$ relating magnetic fields to childhood leukemia of 3.51 (Wald 95% confidence interval: 0.80, 15.4). Next are results from the 3 approximate Bayesian methods described above.

From MCMC sampling, the antilog of the posterior mean of the log odds ratio β_1 is 1.69 (95% posterior interval (PI): 0.58, 4.95). As expected, the odds ratio is shrunk towards the null center of the prior, with approximately 70% of the excess odds eliminated; fortuitously, this posterior mean equals the average odds ratio seen in a pooled analysis of 12 studies, including this study and several much larger ones (12). The precision is dramatically improved. Specifically, the ratio of upper confidence limits to lower confidence limits from maximum likelihood was 19.3, while the ratio of the posterior limits from MCMC is only 8.6. The estimated simulation standard deviation of the MCMC mean of β_1 was 0.0053. From data augmentation, the antilog of the posterior mode of β_1 was 1.74 (95% PI: 0.60, 5.06), similar to the MCMC results. From rejection sampling, the antilog of the posterior mean of the log odds ratio β_1 was 1.69 (95% PI: 0.57, 5.00), which is very close to the MCMC results. The estimated simulation standard deviation of the rejection-sampling mean of β_1 was 0.0034.

The simulation results are further illustrated in Figure 1. Panel A presents 20,000 draws from the joint prior distribution of the log odds ratio β_1 (x-axis) and the intercept β_0 (y-axis). Panel B presents the 20,000 draws from the joint posterior distribution of these parameters obtained by rejection sampling, while panel C presents the 20,000 draws from the joint posterior distribution obtained by the MCMC approach. The rejection-sampling and MCMC displays are indistinguishable, apart from random variability.

Example 2: viral load and incident AIDS

Table 2 presents results from analyses of the HIV cohort data. The maximum-likelihood estimate of the viral-load odds ratio $\exp(\beta_1)$ was 2.92 (95% confidence interval: 1.60, 5.34). In contrast, the geometric mean odds ratio from MCMC sampling was 2.47 (95% PI: 1.41, 4.32); again the odds ratio is shrunk towards the center of the prior. The precision is also improved, but to a lesser extent than in example 1, because the data provide stronger evidence relative to the (same) prior in example 2. The estimated simulation standard deviation of the MCMC mean of β_1 was 0.0006. From data augmentation, the antilog of the posterior mode of β_1 was 2.46 (95% PI: 1.41, 4.29), again close to the MCMC results. For rejection sampling with M constrained at a maximum of 10,000,000, only one-third as many (i.e., 6,718) posterior draws were accepted because the acceptance ratio was exceedingly small, resulting a high rejection rate and a 5.5 times larger simulation error: The estimated simulation standard deviation of the rejection sampling mean of β_1 was 0.0033 (which is still quite small, however). The antilog of the posterior mean of the log odds ratio β_1 was 2.48 (95% PI: 1.41, 4.36), which, as with example 1, is close to the MCMC results.

DISCUSSION

We have illustrated several methods for obtaining Bayesian posterior distributions from commercial software. The methods gave very similar answers in the 2 examples we present. Each method has strengths and limitations.

Data augmentation using offsets is the most computationally rapid procedure, running in about the same time
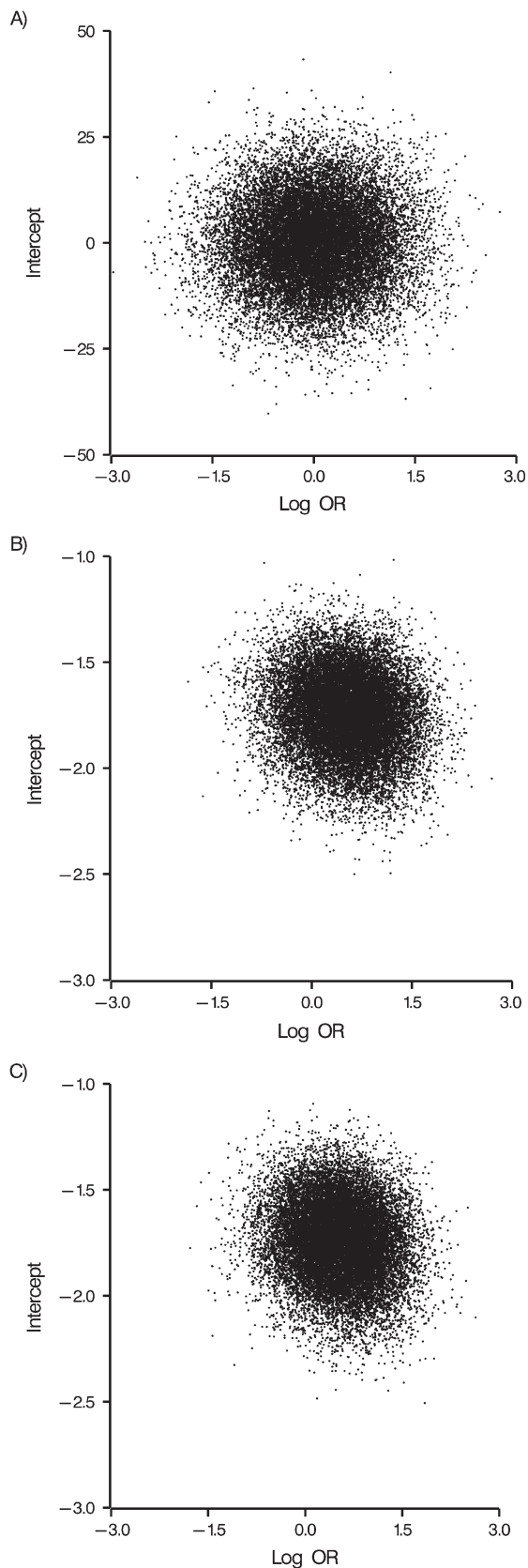
A)

B)

C)

Figure 1. Log odds ratio (OR) for childhood leukemia according to residential exposure to magnetic fields, Denver, Colorado, 1976–1983.

as maximum likelihood; furthermore, its data representation of the prior provides a gauge of the strength of the prior (8) and can be used with any software by modifying the original data set (9). Nonetheless, we have focused on rejection sampling because it may be the most unfamiliar in epidemiology yet has its own advantages. Unlike MCMC methods, rejection sampling generates independent draws from the posterior distribution and raises no concern about convergence to a stationary distribution, which can be a serious issue for MCMC methods (19). This is largely because MCMC draws are serially correlated, whereas rejection sampling produces independent draws. Unlike maximum likelihood and data augmentation but like MCMC methods, rejection sampling need not rely on asymptotic approximations but can instead directly simulate desired exact statistics, such as posterior means and percentiles; for this purpose, it is limited in accuracy only by the computing time required to draw a sufficiently large sample. For simple problems the rejection sampling procedure is faster than MCMC, and the theory behind it is much more transparent (7). In example 1, the MCMC procedure with lag 5 took about 5 minutes to complete, while rejection sampling took about 3 seconds on the same laptop (i.e., a 2.8-GHz dual-core processor, 3 GB of RAM, and SAS version 9.2 under Windows XP). This makes rejection sampling more practical for large-scale sensitivity analyses, in which the data model and prior are varied in a systematic fashion over the joint range of possibilities.

A major limitation of rejection sampling is that the rejection rate increases rapidly with the number of parameters and may lead to unacceptably long run times for models with large numbers of parameters. In example 2, the MCMC with lag 5 took about 14 minutes to yield 20,000 draws, while our rejection sampling approach took over 7 hours to yield one-third as many draws. We expect this limitation of rejection sampling to occur whenever the prior and the likelihood are highly disparate, as when the likelihood is far more informative (i.e., more concentrated) than the prior. One common setting in which this occurs is when highly dispersed (''noninformative'') priors are used. Part of the profound slowdown we observed, however, reflects our use of SAS rather than an efficient compiler language (e.g., C++, Gauss). Other forms of rejection sampling based on taking draws from a distribution approximating the posterior (2, 6) can have dramatically lower rejection rates and hence improved run times but require more sophisticated implementation.

All of the methods we have illustrated require that the modeling problem have a well-behaved likelihood function

Table 2. Estimated Odds Ratios for Incident AIDS According to Human Immunodeficiency Virus Viral Load Set Point, Multicenter AIDS Cohort Study, 1984–1998[a]

Method	Estimate of β_1	SE or SD[b]	OR = exp(β_1)	95% CI or 95% PI[c]
Maximum likelihood	1.072	0.308[d]	2.92	1.60, 5.34[e]
Bayesian[f]				
MCMC[g]	0.905	0.285	2.47	1.41, 4.32
	0.905		2.47	1.42, 4.32
Rejection sampling[g]	0.908	0.288	2.48	1.41, 4.36
	0.900		2.46	1.41, 4.42
Data augmentation[h]	0.901	0.283	2.46	1.41, 4.29

Abbreviations: AIDS, acquired immunodeficiency syndrome; CI, confidence interval; MCMC, Markov chain Monte Carlo; OR, odds ratio; PI, posterior interval; SD, standard deviation; SE, standard error.

[a] Data were obtained from the Multicenter AIDS Cohort Study (13). Results were adjusted for age and race.

[b] Standard deviation unless otherwise specified.

[c] 95% posterior interval unless otherwise specified.

[d] Standard error.

[e] 95% confidence interval.

[f] All Bayesian methods used a lognormal odds-ratio prior with 95% limits of 1/4 and 4.

[g] First estimate of β_1 is the mean of 20,000 draws or 6,718 rejection-sampling draws; first limits are exp(mean $\pm$ 1.96 $\times$ SD), where the mean and SD are computed from the draws. Second estimate is the median; second limits are the 2.5th and 97.5th percentiles. The simulation error was 0.0006 for MCMC and 0.0033 for rejection sampling.

[h] Estimate is the posterior mode (maximum of posterior); limits are exp(mean $\pm$ 1.96 $\times$ SD), where the SD is the β_1 diagonal entry from the inverse of the total (i.e., data and prior) information matrix.

or posterior distribution, something which is typically but not always the case. For example, ordinary maximum-likelihood methods and their variants (i.e., conditional and partial likelihood) assume there is a single and finite maximum of the likelihood function, and data augmentation assumes there is a single and finite maximum of the posterior. MCMC methods require careful diagnostics to assess mixing and convergence, and in some cases all of these diagnostics can fail (19).

Analogously, rejection sampling requires that the distribution used for sampling adequately covers the posterior distribution, which can be hard to judge when there are many model parameters, although comparison of the sampled distribution to the maximum-likelihood and data-augmentation results can help detect problems. In addition, rejection sampling as implemented here also requires a single and finite maximum of the likelihood function. However, large disparities between a prior distribution and the observed likelihood (e.g., a large z score for the difference between the prior mean and the maximum-likelihood estimate) suggest that it is inadvisable to proceed with a Bayesian analysis using that prior (8).

There are other approaches to obtaining posterior parameter distributions that we did not cover—for instance, Laplace's method (6) and approximate Bayesian computation (20). Nonetheless, no matter what method one chooses to compute Bayesian results, it can be valuable to convert the proposed prior into unrescaled data as a measure of how much information the prior is contributing to the final results (8). Unrescaled data augmentation priors provide insight into the strength of the prior that is often lacking when priors are specified only by the parameters that govern the prior distribution. This insight is provided by having to confront a data set that encodes all prior information; the amount of information in these pseudodata (represented by the number of prior cases in typical epidemiologic settings) is difficult to ignore, and the chances of dramatically overstating or understating prior evidence are probably decreased. From there it is a minor step to compute the data augmentation posterior, which can serve as a check on other methods insofar as large disparities among results may indicate a highly nonnormal posterior, an MCMC convergence problem, or a programming error.

Bayesian methods are helpful for handling complex problems in epidemiology, such as sparse data (9), bias analysis (21), and problems with multiple, highly correlated exposures (22). While MCMC procedures have been developed in widely used statistical software packages, such as the SAS procedure MCMC and R-callable WinBUGS, little has been published in the epidemiologic literature to explain how these complex procedures operate or the cautions needed in their use. In future work, we hope to offer a detailed description of MCMC methods for the practicing epidemiologist.

Our rejection-sampling approach may be seen as a bridge to understanding these more complex methods for sampling from posterior distributions, rather than as a general Bayesian method. Nonetheless, rejection sampling can be used as a primary method when the analysis model is relatively simple, as in our examples. Smith and Gelfand's landmark primer on rejection sampling was titled ''Bayesian statistics without tears'' (10). We hope that our review here likewise promotes use of Bayesian methods in epidemiology without unnecessary suffering.

ACKNOWLEDGMENTS

REFERENCES

1. Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. *J Am Stat Assoc*. 1987; 82(398):528–540.
2. Gelfand AE, Smith AFM. Sampling based approaches to calculating marginal densities. *J Am Stat Assoc*. 1990;85(410): 398–409.
3. Berry DA, Stangl D. *Bayesian Biostatistics*. Boca Raton, FL: CRC Press; 1996.
4. Ibrahim JG, Chen MH, Sinha D. *Bayesian Survival Analysis*. New York, NY: Springer Publishing Company; 2001.
5. Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. New York, NY: Chapman & Hall, Inc; 2003.
6. Gelman A, Carlin JB, Stern HS, et al. *Bayesian Data Analysis*. 2nd ed. Boca Raton, FL: CRC Press; 2003.
7. Bedrick EJ, Christensen R, Johnson W. A new perspective on priors for generalized linear models. *J Am Stat Assoc*. 1996; 91(436):1450–1460.
8. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol*. 2006;35(3):765–775.
9. Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *Int J Epidemiol*. 2007;36(1):195–202.
10. Smith AFM, Gelfand AE. Bayesian statistics without tears: a sampling-resampling perspective. *Am Stat*. 1992;46(2):84–88.
11. Savitz DA, Wachtel H, Barnes FA, et al. Case-control study of childhood cancer and exposure to 60-Hz magnetic fields. *Am J Epidemiol*. 1988;128(1):21–38.
12. Greenland S, Sheppard AR, Kaune WT, et al. A pooled analysis of magnetic fields, wire codes, and childhood leukemia. Childhood Leukemia-EMF Study Group. *Epidemiology*. 2000; 11(6):624–634.
13. Kaslow RA, Ostrow DG, Detels R, et al. The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol*. 1987; 126(2):310–318.
14. Silverberg MJ, Wegner SA, Milazzo MJ, et al. Effectiveness of highly-active antiretroviral therapy by race/ethnicity. Tri-Service AIDS Clinical Consortium Natural History Study Group. *AIDS*. 2006;20(11):1531–1538.
15. Muñoz A, Xu J. Models for the incubation of AIDS and variations according to age and period. *Stat Med*. 1996; 15(21-22):2459–2473.
16. Sterne JA, Hernán MA, Ledergerber B, et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. Swiss HIV Cohort Study. *Lancet*. 2005;366(9483):378–384.
17. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457–472.
18. Greenland S. Generalized conjugate priors for Bayesian analysis of risk and survival regressions. *Biometrics*. 2003; 59(1):92–99.
19. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. *J Am Stat Assoc*. 1996;91(434):883–904.
20. Marjoram P, Molitor J, Plagnol V, et al. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A*. 2003; 100(26):15324–15328.
21. Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. *Int J Epidemiol*. 2009;38(6):1662–1673.
22. MacLehose RF, Dunson DB, Herring AH, et al. Bayesian methods for highly correlated exposure data. *Epidemiology*. 2007;18(2):199–207.

APPENDIX 1

Details of the Markov chain Monte Carlo (MCMC) approach

The SAS procedure GENMOD uses a Gibbs sampler based on an adaptive Metropolis rejection sampling algorithm, also known as an ARMS algorithm, to draw a sample from a full conditional distribution. (This approach is described at http://www.maths.leeds.ac.uk/~wally.gilks/adaptive.rejection/web_page/Welcome.html.) A Markov chain is a stochastic process whereby only the current status (and not the past history) of the process is required to make predictions about the future status of the process at a given transition step. A stationary distribution is the limit of the k-step transition intensity, as k goes to infinity.

SAS code for obtaining MCMC posterior samples

```
*Bayes by MCMC;
data prior;
   input _type_ $ Intercept x;
   cards;
   Var 100 0.5
   Mean 0 0
;
*data a contains the original data with y a case indicator, n always 1, and x an exposure indicator;
proc genmod data=a;
   model y/n=x/d=b;
   bayes seed=1 nbi=1000 nmc=100000 thin=5 coeffprior=normal(input=prior) diag=autocorr diag=gelman(nchain=3);
   title "Bayes by MCMC";
```

SAS code for obtaining posterior approximation by data augmentation

```
*Bayes by data augmentation;
data priorint;
   y=500; n=1000; s2=100; s=sqrt(s2*y/2); int=1/s; h=log(y/(n-y))-(log(1)/s); x=0;
data priorx;
   y=500; n=1000; s2=.5; s=sqrt(s2*y/2); x=1/s; h=log(y/(n-y))-(log(1)/s); int=0;
*data a contains the original data with y a case indicator, n always 1, and x an exposure indicator;
data da;
   set a priorint priorx;
   if h=. then h=0;
proc genmod data=da;
   model y/n=int x/d=b noint offset=h;
   title "DA";
```

APPENDIX 2

Example SAS Code for Rejection-Sampling Analysis of Case-Control Data on Residential Magnetic Fields and Childhood Leukemia

```
data post;
   retain count draw 0;
   call streaminit(3);
   rho=0;
   do while (count<20000 and draw<10000000);
      draw=draw+1;
      *create prior;
      intercept=rand("normal")*sqrt(100)+0;
      x=rand("normal")*sqrt(.5*(1-rho**2))+(0+rho*(sqrt(.5)/sqrt(100))*(intercept-0));
      u=rand("uniform");
      *calculate posterior;
      muXeq0=exp(intercept)/(1+exp(intercept));
      muXeq1=exp(intercept+x)/(1+exp(intercept+x));
      if muXeq0=1 then muXeq0=0.99999;
      if muXeq1=1 then muXeq1=0.99999;
      *below are hard-coded numbers from the 2x2 table and maxlogl;
      logl=33*log(muXeq0)+(226-33)*log(1-muXeq0)+3*log(muXeq1)+(8-3)*log(1-muXeq1);
      maxlogl=-99.2495;
      if u<=exp(logl-maxlogl) then do;
         count=count+1;
         output;
      end;
   end;
   keep draw intercept x;
run;
```