Applications of Bayesian Methods to Epidemiologic Research

Richard F. MacLehose · Ghassan B. Hamra

Abstract In recent years, Bayesian methods have been used more frequently in epidemiologic research, perhaps because they can provide researchers with gains in performance of statistical estimation by incorporating prior information. We discuss some of the more common types of Bayesian models in the epidemiologic literature including subjective priors for parameters of interest, weakly informative priors, and hierarchical Bayesian priors. We suggest potential avenues for future research and methods for implementation of Bayesian techniques.

Keywords Bayes · Hierarchical · Multilevel · Shrinkage · Mean squared error · Data augmentation · MCMC

Introduction

Two forms of statistical inference dominate the field of statistics: frequentist and Bayesian. Frequentist estimation emphasizes long-run properties of estimators, such as their bias and mean squared error (MSE). Bayesian estimation, conversely, emphasizes making probability statements about the parameters in statistical models, given some data. For example, Bayesian estimation allows one to make statements regarding the probability that an effect lies between two points, such as the 2.5th and 97.5th values of a parameter's distribution. The two modes of estimation answer different, but complementary, questions; a Bayesian estimator making a statement about the effect size being between two points should be interested

R. F. MacLehose (✉)
Department of Epidemiology and Community Health, University of Minnesota, 1300 S 2nd St #300, Minneapolis, MN 55454, USA
e-mail: macl0029@umn.edu

G. B. Hamra
International Agency for Research on Cancer, Lyon, France

in the long-run frequency of being correct when making such a statement.

Epidemiology, which grew as a discipline alongside the development of frequentist theory and methods, has largely adopted frequentist inferential methods. Despite notable exceptions [1], the vast majority of early statistical analysis in epidemiology was frequentist. Only in recent years has explicit Bayesian analysis become more widespread in epidemiology [2]. We believe this shift has occurred largely because of the appealing properties inherent in Bayesian estimators. Crudely, Bayesian estimators can often be viewed as weighted averages between a prior distribution and the estimate one would obtain from only examining the data at hand [3–5]. That is, the estimate one would obtain from the data alone is moved, or shrunk, in a direction influenced by the prior distribution. This results in some appealing statistical advantages for Bayesian methods. In this article, we review shrinkage estimation and discuss two forms commonly seen in the epidemiologic literature: (1) simple shrinkage towards a fixed point and (2) shrinkage towards nearby estimates via hierarchical regression. Further, we highlight techniques for implementation and offer possible fruitful directions for future application of Bayesian analysis in epidemiology.

Review of Shrinkage Estimation

Common frequentist methods (particularly for regression models) require a researcher to specify a model for the likelihood of the data, depending on parameters of interest. Under assumptions of no systematic error (such as those due to uncontrolled confounding, selection bias, information bias, or model misspecification) and with large sample sizes, maximum likelihood estimators (MLEs) are asymptotically normally distributed and unbiased. For example, an estimator, $\widehat{\beta}$,

from a regression model has a normal sampling distribution centered on the true parameter estimate, β, with variance, se^2.

$$\widehat{\beta} \sim N\left(\beta, se^2\right) \qquad (1)$$

Common frequentist point estimates and confidence intervals typically follow directly from this expression. Bayesian estimation, conversely, is based on the posterior distribution; that is, the distribution of β conditional on the observed data and the prior distribution. This requires a researcher to specify a model for the likelihood of the data collected from their study as well as a prior distribution for unknown parameters. A subjective Bayesian would place a prior distribution on β in equation (1) to express his/her uncertainty in the true parameter value, before conducting the study:

$$\beta \sim N\left(\beta_0, \tau^2\right) \qquad (2)$$

The normal distribution is a computationally convenient choice, though not the only one that could be used. The prior distribution for β is centered on β_0, which would be a subjective Bayesian's "best guess" at the true effect size and τ^2, the prior variance, is a measure of the researcher's certainty in that "best guess." Prior 95 % intervals can be calculated as $\beta_0 \pm 1.96\,\tau$; these prior intervals represent a range of values within which the researcher is 95 % certain the true effect lies, prior to collecting new data.

Bayesian estimation is based on the posterior distribution, which is obtained by combining the likelihood with the prior distribution [5, 6•, 7–9]. In simple well-identified models such as those laid out in equations (1) and (2), with reasonable sample sizes, a posterior effect estimate, θ, is approximately a weighted average of the frequentist effect estimate and the mean of the prior distribution [8]. The weights are a function of the respective variances:

$$\theta = \widehat{\beta}\,\frac{1/se^2}{1/\tau^2 + 1/se^2} + \beta_0\,\frac{1/\tau^2}{1/\tau^2 + 1/se^2} \qquad (3)$$

As the variance in the prior distribution (2) decreases relative to the variance in the sampling distribution (1), the weight attached to the second term in the summation increases and the posterior effect increasingly shrinks away from the MLE and towards the prior mean. If the sampling distribution variance is far smaller than the prior variance ($se^2 \ll \tau^2$), the posterior estimate will approximate the frequentist point

estimate. Conversely, if the prior variance is far smaller than the sampling distribution variance ($se^2 \gg \tau^2$), the posterior estimate will approximate the prior mean.

Bayesian estimation, in this simple setting, can usefully be thought of as akin to a meta-analysis of two studies. The mean and 95 % interval estimate from the prior distribution can be viewed as the summary estimates from the first study in the meta-analysis, while the obtained frequentist estimate and 95 % interval estimate are the second study. The weights used to combine the prior mean and study result in expression (3) are, indeed, equivalent to the inverse variance weights often used in a fixed-effects meta-analysis [10]. The meta-estimate obtained from using inverse-variance weights can serve as a very good approximation to the Bayesian posterior distribution [9].

That the posterior distribution is a weighted average of the prior and likelihood is a natural, and appealing, consequence of Bayesian estimation. However, shrinkage also has desirable frequentist properties and non-Bayesians are also enthusiastic proponents of shrinkage estimators. In the frequentist literature, methods that induce shrinkage are referred to as penalized likelihood methods [11–13]. There is generally a perfect correspondence between frequentist penalized likelihood methods and Bayesian methods: the penalty placed on the likelihood is proportional to the prior distribution. For instance, the penalized likelihood approach known as ridge regression arises from using a normal prior as in (2) with $\beta_0 = 0$. The adoption of these models in traditional frequentist estimation results from the fact that, in a broad range of applications, shrinkage estimators enjoy desirable frequentist properties, such as reduced MSE [5, 14–16].

The simple form of the sampling distribution in (1) and the prior in (2) allow us to easily examine the long-run frequentist characteristics of the Bayesian estimator in (3). Before proceeding to the MSE of the estimate, we consider the amount of bias (the difference between the expected value of this estimator and the truth), if any, introduced by using this Bayesian estimator:

$$Bias(\theta) = (\beta_0 - \beta)\,\frac{1/\tau^2}{1/\tau^2 + 1/se^2} \qquad (4)$$

In the event that the prior mean, β_0, is chosen such that it is exactly equal to the true parameter value, the Bayesian estimator will be unbiased. Alternatively, if the prior variance is infinite, the bias will be zero. This type of prior distribution with infinite variance is referred to as an improper prior; in many simple models, a Bayesian estimator based on an improper prior such as this is equivalent to the frequentist MLE estimator. More generally, however, the bias in (4) will depend

on the relative magnitude of se^2 and τ^2, as well as how far the prior mean is from the truth.

If bias were our sole concern, the Bayesian estimator would seem wholly deficient. However, because Bayesian estimators incorporate prior information, they exhibit a decreased variance relative to traditional frequentist estimators. The MSE of an estimator is a measure of the variability of this estimator around the truth. MSE can be calculated as the sum of the square of the bias and the variance of the estimator:

$$MSE(\theta) = \left[(\beta_0 - \beta) \frac{1/\tau^2}{1/\tau^2 + 1/se^2} \right]^2 + \frac{1/se^2}{\left(1/\tau^2 + 1/se^2 \right)^2} \quad (5)$$

Because MSE incorporates both bias and variance, it is possible for an estimator to be biased with low variance and thus have lower MSE than an unbiased estimator with high variance. Indeed, the Bayesian estimator in (3) can be just such an estimator.

We consider a hypothetical example to illustrate the potential for improved frequentist properties when using Bayesian estimators. Suppose three researchers (A, B, and C) are interested in estimating the effect of cumulative occupational asbestos exposure on the risk of mesothelioma. The researchers plan a study and are able to identically repeat the study a large number of times. Further, the true relative risk (RR), unbeknownst to the researchers, is $\beta = \log(RR) = \log(2)$ with $se^2 = 0.1$. Researcher A will analyze each study using frequentist methods associated with equation (1), while researchers B and C will use the Bayesian estimator in equation (3). Both Bayesian researchers use substantive knowledge to specify the prior mean as $\beta_0 = \log(1.2)$ which, we note, is a poor estimate of the truth. Researcher B specifies the prior variance as $\tau^2 = 0.01$ (implying prior 95 % interval for the RR: 1.0–1.5) and researcher C specifies $\tau^2 = 0.5$ (implying prior 95 % interval for the RR: 0.3–4.8). Each researcher analyzes each data set after each replication of the study. After the large number of replications is complete, the researchers compare their accumulated effect estimates. Researcher A, the frequentist, finds that the average of his/her estimated effects is equal to $\log(2)$, and therefore unbiased. Researcher B, the Bayesian with a more precise prior, has estimated effects that have a bias (on the log scale) of −0.46. Researcher C, the Bayesian with a less precise prior, has estimated effects that have a bias (on the log scale) of −0.09. Both Bayesian researchers have produced biased effects; their estimates have all been shrunk back to the incorrectly specified prior mean. Researcher B, who had the most precise prior specification, induced the most shrinkage and therefore had the most bias. However, when the researchers compare their estimates in terms of MSE, researcher A has MSE=0.1, researcher B has MSE=0.22, and researcher C has MSE=0.08. Despite producing estimates that

were biased on average, researcher C is also closest to the truth on average. Indeed, Bayesian estimators often have decreased MSE, compared with comparable frequentist estimators, when the prior mean is "close" to the truth. How close is close enough to guarantee that a Bayesian estimator has lower MSE than a frequentist estimator is difficult to answer in realistic settings. However, Bayesian implementations with greater prior variance are more robust to the misspecification of the prior mean. Hence, the general practical advice is to err on the side of greater prior variance [9].

Shrinkage Towards a Fixed Point

Because Bayesian shrinkage estimators can enjoy reduced MSE when the prior mean is reasonably close to the truth, many authors have recommended Bayesian analyses as a way to improve model performance [4, 17, 18•]. A common use of Bayesian models in epidemiology is to use a vague prior distribution to shrink effect estimates towards a fixed point: the prior mean. Vague priors typically set $\beta_0 = 0$ in expression (2) so that all shrinkage is towards the null. This approach may be appealing to researchers attempting to limit the number of false-positive results, and is often viewed as a conservative approach. What prior variance constitutes a vague prior is necessarily a subjective consideration. A prior variance that specifies a suitably vague prior in cardiovascular epidemiology may be highly informative for infectious disease epidemiology. Gelman et al. suggest a default shrinkage prior but replace the normal distribution in (2) with a heavier-tailed distribution that shrinks large effects to a lesser extent [17]. Hamra et al. suggest using a vague prior, $N(\beta_0 = 0, \tau^2 = 1.33)$, as a sensitivity analysis. In their approach, the discrepancy between the frequentist results and the results using the vague prior suggest the extent to which the results are dependent on the relatively small amount of additional information incorporated via the prior distribution.[18•].

When vague priors have been implemented in epidemiologic research, the most common occurrence seems to be in settings in which investigators attempt to simultaneously estimate many effects. Herdt-Losavio shrunk the effect of various maternal occupations on risk of birth defects using $\tau^2 = 1.125$ [19]. D'Aloisio (2009) used shrinkage methods to stabilize estimates of genetic polymorphisms on insulin growth factor levels and chose their prior variance based on the variability of their continuous outcome [20]. Similarly, O'Brien (2014) recently implemented vague priors in the analysis of a large study of breast cancer susceptibility loci [21].

Because choosing a prior mean that is close to the truth decreases MSE, an easy way to improve on vague priors is to estimate the prior mean from previous literature, if any exists. This use of the prior literature depends not only on the availability of such literature but also on its quality. Specifying a

prior mean using poor prior data could easily increase the bias and MSE of the resulting estimator. However, despite repeated efforts to encourage the use of such subjective priors by many authors [8, 9, 22, 23], there are very few examples of prior distributions explicitly informed by prior studies in the epidemiologic literature [19, 24].

Hierarchical Shrinkage

The shrinkage estimators described above improve MSE by directly incorporating prior knowledge about expected or previously reported effects. Results are shrunk towards a fixed point; either the null or a point defined by previous research. In other cases, such as when a researcher is simultaneously estimating multiple effects, extensions of the prior distribution in (2) can incorporate substantive knowledge so that effect estimates shrink toward a point estimated, in part, from the data [4, 25]. Bayesian estimation in this setting allows effect estimates to be shrunk towards a function of other estimates in the model; for instance, effects in a similar exposure group might all be shrunk towards one another.

For example, Engel et al [26]. examined the effect of polymorphisms in multiple inflammatory cytokines on small-for-gestational-age birth. Sparse data and correlation resulted in unstable estimates when all effects were included simultaneously in the model. Instead, the authors assumed a prior distribution in which the effects of the polymorphisms in proinflammatory cytokines were assumed to be random samples from one distribution and the effects of polymorphisms in anti-inflammatory cytokines were assumed to be random samples from a different distribution; this is often referred to as hierarchical Bayesian modeling. This type of hierarchical model assumes the polymorphism effects in each group are exchangeable. It allows the shrinkage of each individual effect towards the grand mean of all individual effects in that group. This prior structure allows for improved performance by leveraging additional information: the estimate of each effect "borrows" information from the other effect estimates in that group through the use of the prior distribution. In the Engel example, all proinflammatory cytokine, polymorphism effect estimates were shrunk back to the grand mean of all the individual estimates. This borrowing of information and shrinkage towards the group mean can result in substantial reduction in MSE. This type of hierarchical model can be implemented in a semi-Bayes (in which prior distributions are specified for some terms in the model) [4, 15, 27, 28] or fully-Bayes (in which prior distributions are specified for all terms in the model) [3] framework. A growing number of researchers are using these hierarchical methods, including Hamra et al [29]. who conducted a similar analysis to examine multiple, highly correlated, asbestos fiber size groups and

O'Brien et al. who examined breast cancer susceptibility loci [30••].

Proper grouping of effect estimates and hierarchical structure is a substantive concern that is dictated by the application. In spatial epidemiology, it is common to estimate effects (such as standardized mortality ratios) in geographic regions such as states, counties, or even census tracts [31–34]. The mutually exclusive groupings in the previous examples are relaxed in this case. Conditional autoregressive priors are commonly used to allow effects from one geographic region to be shrunk towards the effect of neighboring geographic regions. Similar autoregressive priors are also used in time-series data to allow the estimated effect at one point in time to be shrunk towards effects at adjacent time points. Other researchers have attempted to remain agnostic about which parameters should shrink towards one another and allow the data to determine the groupings of coefficients [3]. Careful consideration should be given to the specification of the hierarchical structure as results may be sensitive to these assumptions.

Bayesian Implementation

Bayesian methods are often more difficult to implement than traditional frequentist methods, a fact that has hindered their adoption in the epidemiologic literature [8]. Bayesian models are often fit using Markov Chain Monte Carlo (MCMC) techniques. While implementing these techniques has become easier in recent years, they still require special attention to technical details in which epidemiologists do not typically receive training.

Data Augmentation

In response to technical difficulties in implementing MCMC procedures, numerous authors have proposed data augmentation procedures; [8, 9, 23, 35–37, 38••] this approach appends hypothetical data representing the prior distribution to the actual data and relies on standard model fitting techniques to estimate Bayesian posterior distributions. Greenland [9] gives details on translating a prior distribution into a dataset as well as methods for improving the accuracy of approximations. Standard regression models can be run on the combined data, with an indicator variable designating whether the observations are derived from the actual or prior data. The regression estimate and confidence intervals for the exposure of interest can be treated as the mean and credible intervals (the Bayesian analog of the confidence interval) of the Bayesian posterior distribution.

Software Packages for Markov Chain Monte Carlo

Data augmentation serves as an easy way to obtain an approximation to the posterior distribution that is often sufficiently accurate. There may be instances, however, when MCMC is preferable because it provides the opportunity to estimate the exact posterior distribution. Further, researchers often use Bayesian methods to handle complex analytic problems; these may be regression models with complicated forms, or situations in which a researcher would like to specify a non-standard or even non-parametric prior. For example, a user may desire to constrain parameters based on a priori knowledge of their rank ordering [39], or use an inequality constraint to adjust for RR estimates that exceed a probability of 1.0 [40]. In these settings, data augmentation may be difficult to implement, and a researcher may need to implement a MCMC estimation procedure, which has been described in detail elsewhere [41, 42]. MCMC is implemented with varying degrees of difficulty and using different sampling procedures in a few statistical software packages. We will discuss MCMC as implemented in Statistical Analysis Software (SAS), JAGS, WinBUGS, and STAN.

SAS

SAS provides two mechanisms for using MCMC. First, a user can specify the BAYES statement within the packaged procedures GENMOD, PHREG, and LIFEREG; this statement is a single extra line of software code that invokes MCMC rather than maximum likelihood estimation. This allows epidemiologists relatively easy implementation of Bayesian linear, logistic, Poisson, and Cox models. These procedures are a convenient option for including a prior distribution for model parameters of interest. Second, there is an independent MCMC procedure, which allows the user greater flexibility in specifying many prior distributions and likelihood functions. A user can specify complex hierarchical regression models, whereas the BAYES statement allows relatively simple prior specifications. When possible, SAS will use a conjugate sampling procedure, which is statistically efficient. Otherwise, the most common sampling algorithms used by SAS are the adaptive rejection Metropolis or random walk Metropolis algorithms. Sullivan and Greenland provide an introduction to Bayesian regression using SAS [38••].

JAGS and WinBUGS

Unlike SAS, JAGS and WinBUGS are purpose-built software packages; they are designed to implement Bayesian models with the Gibbs Sampler. JAGS is an acronym for Just Another Gibbs Sampler [43] and WinBUGS stands for Bayesian inference Using Gibbs Sampling [44]. WinBUGS is a general Windows-based user interface program that allows specifying models, sampling from the posterior distribution of parameters of interest, diagnosing model convergence, and creating graphical and analytic output. JAGS, which uses a similar language for model specification as WinBUGS, allows the user to specify a model and generate samples from the posterior distribution; however, this is where the features of JAGS end. Thus, an individual using JAGS should be familiar with the R software package to analyze the output and conduct tests for model convergence. Packages such as rjags and mcmcplots are available and allow users to create graphics and analyze output from JAGS; both WinBUGS and JAGS can be invoked using R statistical software. Many textbooks offer detailed descriptions for using WinBUGS [45, 46].

STAN

STAN is named after the co-inventor of the Monte Carlo method, Stanislaw Ulam, and is unlike the SAS, JAGS, or WinBUGS samplers [47]. STAN uses a Hamiltonian Monte Carlo and No-U Turn sampling procedure, which its creators chose because of its ability to handle problems with which WinBUGS and JAGS may have trouble; namely, use of non-conjugate priors and high posterior correlations. While the language for STAN is similar to that of WinBUGS and JAGS, it is not identical. As with JAGS, a user can specify and call a STAN model using an R package, namely rstan. Additionally, the resulting output can be analyzed with the same packages used to analyze the output from JAGS.

Discussion

Bayesian methods can provide researchers with gains in performance of statistical estimation by incorporating prior information. We have discussed a few of the common forms the prior may take including subjective values for parameters of interest, weakly informative default priors, and hierarchical Bayesian priors. Examples of researchers successfully implementing Bayesian methods to improve their analyses were also provided.

We have two recommendations to further the use of Bayesian methods in epidemiologic research. First, the lack of articles in which the prior distributions are based on previous research is notable. The vast majority of articles that employ simple shrinkage techniques choose to shrink estimates towards the null, rather than towards some substantively reasonable location. Shrinkage towards the null offers researchers the opportunity to justify their approach with claims of it being conservative. However, shrinkage towards locations based on prior knowledge may lead to better estimators. Employing

shrinkage towards substantively justified locations will re-quire an extensive examination of the literature or consultation with substantive experts. The strengths and limitations of existing literature should be carefully considered and studies that are not of sufficient quality should be excluded from consideration for informing priors. If multiple studies are available for inclusion in a prior distribution, meta-analyses can be used to aggregate them.

Second, the data augmentation approach has seen relatively littleuse in the literature despite having been introduced in epidemiology in 2007 [9]. This approach is easy to implement and brings Bayesian results within the grasp of any epidemi-ologist who can run standard regression models. Data aug-mentation techniques cleverly avoid the need for epidemiolo-gists to learn specialized computing techniques such as MCMC. While data augmentation may be difficult to extend to more complicated hierarchical structures, its ease of imple-mentation and transparency for simple shrinkage is laudable. We encourage analysts to explore these techniques.

While we have restricted our focus to shrinkage techniques, there is another common use of subjective Bayesian analysis: when prior information, informed by previous studies or ex-pert opinion, is used in bias analysis [48–51]. Adjustments for possible biases in epidemiologic studies are dependent on bias parameters, such as the sensitivity or specificity of classifica-tion of the exposure. Informative values chosen for these parameters would, ideally, be based on validation sub-studies. However, lacking these they can be based on relevant prior literature, such as previous validation studies that esti-mated the sensitivity and specificity of the exposure measure-ment. Distributions for these bias parameters represent the uncertainty surrounding the bias parameter and as has been noted, there is a correspondence between bias analysis proce-dures and Bayesian procedures [48, 52]. In fact, the bias parameter distribution is simply a Bayesian prior distribution. However, much of the bias analysis literature has avoided the Bayesian foundation of the procedure and focused on practical implementation instead.

Conclusion

We have highlighted many situations in which incorporat-ing even a small amount of prior information can result in Bayesian estimators that have excellent frequentist proper-ties. Furthermore, approximate Bayesian methods, ad-vances in computational efficiency, and efforts to create applicable software packages have made implementation of Bayesian methods much more accessible for interested researchers. As a result, we believe Bayesian methods should be a vital component of every epidemiologist's toolkit and should be used with regularity.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1.   Cornfield J. Recent methodological contributions to clinical trials. Am J Epidemiol. 1976;104:408–24.
2.   Pearce N, Corbin M. Why we should be Bayesians (and often already are without realizing it). In: *Curent Topics in Occupational Epidemiology.* Oxford University Press; 2013.
3.   MacLehose RF, Dunson DB, Herring AH, Hoppin JA. Bayesian methods for highly correlated exposure data. Epidemiology. 2007;18(2):199–207.
4.   Greenland S, Poole C. Empirical-Bayes and semi-Bayes ap-proaches to occupational and environmental hazard surveillance. Arch Env Heal. 1994;49(1):9–16.
5.   Carlin BP, Louis TA. Bayesian methods for data analysis. London: CRC Press; 2009.
6.• Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. Boca Raton: CRC Press; 2013. p. 675. *An excellent all-encompassing introduction to Bayesian analysis.*
7.   Christensen R, Johnson WO, Branscum AJ, Hanson TE. Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians. Boca Raton: CRC Press; 2011.
8.   Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol. 2006;35(3):765–75.
9.   Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. Int J Epidemiol. 2007;36(1):195–202.
10. Egger M, Davey Smith G, Altman DG. Systematic Reviews in Health Care: Meteggera-Analysis in Context. London: BMJ Publishing Group; 2001.
11. Greenland S. Relaxation Penalties and Priors for Plausible Modeling of Nonidentified Bias Sources. Stat Sci. 2009;24(2): 195–210.
12. Park T, Casella G. The Bayesian lasso. J Am Stat. 2008;103(482): 681–6.
13. Hoerl A, Kennard R. Ridge regression: applications to nonorthogonal problems. Technometrics. 1970;12(1):69–82.
14. Hoff P. A first course in Bayesian statistical methods. Dordrecht: Springer; 2009.
15. Greenland S. Principles of multilevel modelling. Int J Epidemiol. 2000;29(1):158–67.
16. Little R. Calibrated Bayes: A Bayes/frequentist roadmap. Am Stat. 2006;60(3):1–11.
17. Gelman A, Jakulin A, Pittau M, Su Y. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat.* 2008:1360–1383.
18.• Hamra GB, MacLehose RF, Cole SR. Sensitivity analyses for sparse-data problems-using weakly informative bayesian priors.

Epidemiology. 2013;24(2):233–9. *This article describes a simple type of vague shrinkage prior. It can be used a sensitivity analysis to detect sparse data.*

19. Herdt-Losavio ML, Lin S, Chapman BR, et al. Maternal occupation and the risk of birth defects: an overview from the National Birth Defects Prevention Study. Occup Environ Med. 2010;67(1):58–66.

20. D'Aloisio AA, Schroeder JC, North KE, et al. IGF-I and IGFBP-3 polymorphisms in relation to circulating levels among African American and Caucasian women. Cancer Epidemiol Biomarkers Prev. 2009;18(3):954–66.

21. O'Brien KM, Cole SR, Engel LS, et al. Breast cancer subtypes and previously established genetic risk factors: a Bayesian approach. Cancer Epidemiol Biomarkers Prev. 2014;23(1):84–97.

22. Dunson DB. Commentary: practical advantages of Bayesian analysis of epidemiologic data. Am J Epidemiol. 2001;153(12):1222–6.

23. Greenland S. Bayesian perspectives for epidemiologic research: III. Bias analysis via missing-data methods. Int J Epidemiol. 2009;38(6):1662–73.

24. Vinikoor LC, Kaufman JS, MacLehose RF, Laraia BA. Effects of racial density and income incongruity on pregnancy outcomes in less segregated communities. Soc Sci Med. 2008;66(2):255–9.

25. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press; 2007.

26. Engel SAM, Erichsen HC, Savitz DA, Thorp J, Chanock SJ, Olshan AF. Risk of Spontaneous Preterm Birth is Associated With Common Proinflammatory Cytokine Polymorphisms. Epidemiology. 2005;16(4):469–77.

27. Witte JS, Greenland S, Kim LL. Software for hierarchical modeling of epidemiologic data. Epidemiology. 1998;9(5):563–6.

28. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. Epidemiology. 1994;5(6):612–21.

29. Hamra GB, Loomis D, Dement J. Examining the association of lung cancer and highly correlated fibre size-specific asbestos exposures with a hierarchical Bayesian model. *Occup Environ Med*. 2014.

30.•• O'Brien KM, Cole SR, Poole C, et al. Replication of breast cancer susceptibility loci in whites and African Americans using a Bayesian approach. Am J Epidemiol. 2014;179(3):382–94. *This is an excellent example of hierarchical modelling. The authors use both semi-Bayes and fully-Bayes approaches.*

31. Lawson AB. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, Second Edition*. CRC Press; 2013.

32. Kramer MR, Williamson R. Multivariate bayesian spatial model of preterm birth and cardiovascular disease among georgia women: Evidence for life course social determinants of health. Spat Spatiotemporal Epidemiol. 2013;6:25–35.

33. Banerjee S, Gelfand AE, Carlin BP. Hierarchical Modeling and Analysis for Spatial Data. Boca Raton: Chapman & Hall; 2004.

34. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons; 2004.

35. Bedrick EJ, Christensen R, Johnson W. Bayesian Binomial Regression: Predicting Survival at a Trauma Center. Am Stat. 1997;51(3):211–8.

36. Bedrick EJ, Christensen R, Johnson W. A New Perspective on Priors for Generalized Linear Models. J Am Stat Assoc. 1996;91(436):1450–60.

37. Greenland S, Christensen R. Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression. Stat Med. 2001;20(16):2421–8.

38.•• Sullivan SG, Greenland S. Bayesian regression in SAS software. Int J Epidemiol. 2013;42(1):308–17. *An excellent guide to estimating Bayesian models using SAS software.*

39. Hamra G, Richardson D, Maclehose R, Wing S. Integrating informative priors from experimental research with Bayesian methods: an example from radiation epidemiology. Epidemiology. 2013;24(1):90–5.

40. Chu H, Cole SR. Estimation of risk ratios in cohort studies with common outcomes: a Bayesian approach. Epidemiology. 2010;21(6):855–62.

41. Gilks W, Richardson S, Spiegelhalter D. Markov Chain Monte Carlo in Practice. Boca Raton: Chapman & Hall/CRC; 1996.

42. Hamra G, MacLehose R, Richardson D. Markov chain Monte Carlo: an introduction for epidemiologists. Int J Epidemiol. 2013;42(2):627–34.

43. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proc hte 3rd Int Work Distrib Stat Comput (DSC 2003)*. 2003:20–22.

44. Lunn D, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat Comput. 2000;10(4):325–37.

45. Congdon P. Applied Bayesian Modelling. Chichester: John Wiley & Sons; 2003.

46. Congdon P. Bayesian Statistical Modelling. Chichester: John Wiley & Sons; 2007.

47. Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv Prepr arXiv 11114246*. 2011.

48. Greenland S. Multiple-bias modelling for analysis of observational data (with discussion). J R Stat Soc Ser A. 2005;168(2):267–306.

49. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity Analysis of Misclassification: A Graphical and a Bayesian Approach. Ann Epidemiol. 2006;16(11):834–41.

50. Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am J Epidemiol. 2004;160(4):384–92.

51. MacLehose RF, Olshan AF, Herring AH, Honein MA, Shaw GM, Romitti PA. Bayesian methods for correcting misclassification: an example from birth defects epidemiology. Epidemiology. 2009;20(1):27–35.

52. MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian? Epidemiology. 2012;23(1):151–8.