

Sensitivity Analyses for Sparse-Data Problems—Using Weakly Informative Bayesian Priors

Ghassan B. Hamra,^a Richard F. MacLehose,^{b,c} and Stephen R. Cole^a

Abstract: Sparse-data problems are common, and approaches are needed to evaluate the sensitivity of parameter estimates based on sparse data. We propose a Bayesian approach that uses weakly informative priors to quantify sensitivity of parameters to sparse data. The weakly informative prior is based on accumulated evidence regarding the expected magnitude of relationships using relative measures of disease association. We illustrate the use of weakly informative priors with an example of the association of lifetime alcohol consumption and head and neck cancer. When data are sparse and the observed information is weak, a weakly informative prior will shrink parameter estimates toward the prior mean. Additionally, the example shows that when data are not sparse and the observed information is not weak, a weakly informative prior is not influential. Advancements in implementation of Markov Chain Monte Carlo simulation make this sensitivity analysis easily accessible to the practicing epidemiologist.

(*Epidemiology* 2013;24: 233–239)

Epidemiologic studies often must cope with sparse-data problems due to small sample sizes or to reduction in effective sample size due to the study of very uncommon (or common) exposures¹ or highly correlated variables.² However, determining the presence and impact of sparse data in a given study is not as clear as one might believe. In studies with only a few categorical variables, the researcher may be able to identify the presence of sparse data by observing the sample size within cells of contingency tables.³ However, as data become more complex, this approach becomes untenable. In a regression model with a large number of covariates,

researchers should be concerned about the potential impact of data sparseness. One way to quantify the impact of sparse data on parameter estimates is with a sensitivity analysis in which the observed data are augmented with a small amount of additional data (for instance, a few additional exposed cases and controls). Quantifying the degree of change in the parameter estimates that results from the addition of a small amount of additional information represents an informal assessment of the impact of sparse data. If the data are sparse, model estimates will be sensitive to this additional information. If the data are not sparse, model estimates will be robust to the added information.

A method for testing the sensitivity of particular model parameters to sparse data is a natural complement to existing methods for evaluating systematic errors due to confounding, measurement error, and selection bias.^{4–7} Augmenting the observed data with a small amount of additional information is easily accomplished with Bayesian analysis. To this end, we propose the use of a weakly informative prior. Recent advancements in Markov Chain Monte Carlo techniques make this approach easy to use in day-to-day regression analyses. We provide two examples (one as an eAppendix, <http://links.lww.com/EDE/A650>) of odds ratios obtained from conditional logistic regression models, which are notoriously susceptible to sparse-data problems.

WEAKLY INFORMATIVE PRIORS

Inclusion of a prior in a regression model is a simple means of representing the body of knowledge for a parameter of interest external to the study that generated the data.^{8,9} The degree of support for this belief is inversely related to the variance of the prior; that is, the smaller the variance, the more support for the prior. In epidemiologic analyses, researchers may have a priori beliefs that large effect estimates of an exposure–outcome relationship are very unlikely. In studies of common exposures or widespread environmental pollutants, this belief is well-founded, as it is exceedingly rare to see relative measures of effect greater than 10 or less than 1/10. Indeed, outside of a few areas such as infectious disease, exposures are unlikely to be highly associated with the outcome.¹⁰ As regression models become more complex, it is increasingly difficult to determine whether large effects are based on a reliable amount of information within the data or result from problems of data sparseness. Standard maximum

Submitted 13 June 2012; accepted 30 November 2012; posted 18 January 2013. From the ^aDepartment of Epidemiology, UNC Chapel Hill, Chapel Hill, NC; ^bDivision of Biostatistics, University of Minnesota, Minneapolis, MN; and ^cDivision of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN.

Supported by the Centers for Disease Control and Prevention (grant number 1R03OH009800-01), National Institute of Environmental Health Sciences (training grant ES07018), and National Institute of Health (grant number 1U01-HD061940).

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com). This content is not peer-reviewed or copy-edited; it is the sole responsibility of the author.

Correspondence: Ghassan Hamra, Department of Epidemiology, CB# 7435, Chapel Hill, NC 27599-7435. E-mail: ghassan.hamra@unc.edu.

Copyright © 2013 by Lippincott Williams & Wilkins

ISSN: 1044-3983/13/2402-0233

DOI: 10.1097/EDE.0b013e318280db1d

TABLE. Association of Lifetime Alcohol Consumption with Head and Neck Cancer

Alcohol Consumption (L)	No. Cases/ No. Controls	Conditional Maximum Likelihood OR (95% CI)	OR Including Weakly Informative Prior (95% PI)	Change in Estimate %
Oropharyngeal cancer				
0	27/280			
>0-133	69/466	0.93 (0.54-1.62)	0.84 (0.53-1.37)	10.2
134-758	94/360	1.48 (0.83-2.64)	1.40 (0.87-2.27)	5.6
759+	120/173	4.49 (2.40-8.39)	3.26 (1.93-5.44)	32.0
Hypopharyngeal cancer				
0	1/280			
>0-133	5/466	2.25 (0.26-19.84)	0.79 (0.25-2.61)	104.7
134-758	9/360	5.13 (0.61-43.04)	1.64 (0.55-4.91)	114.0
759+	36/173	28.74 (3.42-241.40)	8.64 (3.16-26.37)	120.2

PI, Bayesian posterior intervals.

likelihood estimators are not well suited for analyses of sparse or highly correlated data. Maximum likelihood estimation relies on asymptotic theory, which typically guarantees that estimators are unbiased with an infinite sample size. However, with small sample sizes these estimators may be highly biased.³ Despite the fact that conditional maximum likelihood estimators were developed to deal with sparse data in matched case-control studies,^{11,12} they are themselves subject to sparse-data problems, which occur when there are a large number of strata defined by matching factors and limited data within these strata. This limitation may produce unstable parameter estimates.

Other researchers have shown the utility of correcting for sparse-data problems, and have presented techniques such as use of data-augmentation priors.^{13,14} These priors, which are Bayesian in nature, have been applied using maximum likelihood estimators, which rely on asymptotic assumptions.¹⁵ The use of data-augmentation priors requires a rescaling step to improve its asymptotic approximation. This step is unnecessary when implementing a weakly informative prior using Markov chain Monte Carlo.^{3,14,15} Markov Chain Monte Carlo methods can be used to incorporate information external to the data; however, with current versions of Statistical Analysis Software (SAS), implementing Markov Chain Monte Carlo techniques can be easier than traditional data-augmentation approaches.

This advance in statistical software transforms the evaluation of model-estimate-sensitivity into a potentially routine procedure for the practicing epidemiologist. To this end, we propose a generic weakly informative prior based on an a priori expectation of the magnitude of the relation between an exposure and outcome of interest. For general application, we recommend a normally distributed prior for a regression coefficient, β , such that mean $\mu = 0$ and variance $\sigma^2 = 1.38$. In effect, this says that before conducting the analysis, we are 95% certain that the relative effect estimate

is between 0.1 to 10 on a ratio scale, centered at 1.00, or null. To mimic the weight of information contributed by this weakly informative prior, one may think of a data-augmentation approach where the researcher adds three observations to each cell in a 2×2 table: the mean and approximate variance of the log odds ratio (obtained using Woolf's formula) from a 2×2 table in which all cells contain three observations are 0 and 1.33, respectively.^{16,17} This variance is calculated as follows:

$$\text{variance} = \frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{01}} + \frac{1}{n_{00}} = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1.33$$

Keeping this in mind, one way to treat the result of a sensitivity analysis using a weakly informative prior is the parameter estimate that would have resulted had the investigator observed a small amount of additional data that reflected the null hypothesis. In a situation where a wealth of previous evidence supports a harmful (or protective) effect of the risk factor with the outcome, the null-centered prior can be easily adjusted to reflect this knowledge.

A weakly informative prior is a relatively weak statement of prior knowledge and is tenable in most epidemiologic settings.^{18,19} As the sample size of the study increases, a weakly informative prior will have vanishing impact on model estimates. Specifically, as data become less sparse, we would obtain approximately the same point and interval estimates with or without a weakly informative prior. However, in the presence of sparse data, a weakly informative prior will help stabilize estimation and shrink the unstable and potentially biased maximum likelihood estimates toward the prior mean. We present a worked example of calculating odds ratios below, and we provide a second (simpler) example (along with data and SAS code) as an electronic supplement.

EXAMPLE: ALCOHOL CONSUMPTION AND ORAL CANCER

Hakenewerth et al²⁰ studied the relationship of alcohol consumption and oral cancer among the Carolina Head and Neck Cancer Epidemiology Study, a population-based case-control study of squamous cell carcinoma of the head and neck conducted in North Carolina between 2002 and 2006. The authors analyzed data on 1227 cases and 1325 controls who were frequency-matched on age (25–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–80 years), race (European–American, African–American), and sex, creating 28 matched strata. Additionally, the data include information regarding continuous duration of cigarette smoking (as total years of smoking), a known confounder of the relationship alcohol consumption and oral cancer.^{21,22} Because this is a frequency-matched case-control design, the authors conducted conditional logistic regression analyses where the relationship between lifetime alcohol consumption and oral cancer was evaluated, conditional on race, age, and sex and controlling for continuous years of cigarette smoking. Lifetime alcohol consumption, in liters (L), is divided into four ordered categories of exposure (0, >0–133, >133–758, and >758), and cigarette consumption is treated as continuous. We recreate the authors’ original analysis, estimating the odds of head and neck cancer associated with alcohol consumption, adjusting for continuous years of cigarette exposure, age, race, and sex.

STATISTICAL ANALYSIS

We assume the outcome, y_{ik} , for individual $i = 1, 2, \dots, I_k$ in stratum $k = 1, 2, \dots, K$, follows a logistic model with stratum-specific intercepts, α_k

$$\Pr(y_{ik} = 1) = \{1 + \exp(-\alpha_k - \mathbf{x}_{ik}\beta)\}^{-1} \quad (1)$$

where \mathbf{x}_{ik} is a vector of covariates and β is the vector of coefficients of interest. Under the conditional logistic regression approach, estimation of α_k is avoided by specifying the likelihood of the data conditional on each stratum. Because of the frequency-matched data used in this example, the number of cases, n_{1k} , and controls, n_{0k} , in each stratum may vary, and a general form of the conditional likelihood can be specified as:

$$\prod_{k=1}^K \frac{\prod_{i=1}^{I_k} \exp(y_{ik} \mathbf{x}_{ik} \beta)}{\sum_{d_k \in S_k} \prod_{i=1}^{I_k} \exp(d_{ik} \mathbf{x}_{ik} \beta)} \quad (2)$$

where the s_k is all possible combinations of n_{1k} cases and n_{0k} controls in the k th stratum, and d_k is a vector of one of the possible combinations with d_{ik} , an element in d_k .²³

A typical frequentist approach to analyzing matched case-control data would involve maximizing (2) with respect to β . Implementing our weakly informative prior involves a relatively simple extension of the conditional likelihood. The posterior distribution of interest is proportional to the

conditional likelihood in expression (2) times the weakly informative prior. Because our weakly informative prior has the form of independent normal priors for the log odds ratios, our posterior distribution is proportional to

$$\prod_{k=1}^K \frac{\prod_{i=1}^{I_k} \exp(y_{ik} \mathbf{x}_{ik} \beta)}{\sum_{d_k \in S_k} \prod_{i=1}^{I_k} \exp(d_{ik} \mathbf{x}_{ik} \beta)} \prod_j \exp\left(-\frac{\beta_j^2}{2 \times 1.33}\right) \quad (3)$$

where $j = 1$ to J indexes the J regression coefficients.

We note that the general form of expression (3) is the same as frequentist procedures that penalize the likelihood, such as ridge regression.²⁴ Indeed, the weakly informative prior we have specified is a Bayesian analog of ridge regression in which the tuning parameter is specified based on prior knowledge.

We used a Gibbs sampler to run Bayesian models for 10,000 iterations with a burn-in of 1,000 iterations. A Gelman–Rubin diagnostic check was conducted for three chains to confirm convergence of the Markov Chain Monte Carlo procedure. Trace and autocorrelation plots indicate model convergence. When reporting results, we provide 95% confidence intervals and 95% posterior intervals (PIs) for non-Bayesian and Bayesian models, respectively. All analyses were conducted in SAS version 9.2 (SAS Institute, Cary, NC).

RESULTS

The Table provides an example of model results that are moderately and highly sensitive to sparse data. We highlight the percent change in the odds ratios of oropharyngeal and hypopharyngeal cancer for each category of alcohol consumption compared with no alcohol consumption. For oropharyngeal cancer, the conditional maximum likelihood estimates are moderately precise, suggesting that data are not overly sparse across strata of the matching factors. The odds ratios for oropharyngeal cancer associated with low, medium, and high levels of lifetime alcohol consumption, relative to none, change by 10%, 6%, and 32%, respectively, when a weakly informative prior is incorporated. The precision of the estimates is largely unaffected by the weakly informative prior, with the exception of the highest category of alcohol exposure, for which the upper bound of the 95% PI is more strongly attenuated toward the prior mean. Figure 1 shows the relative contribution to the posterior distributions of the odds ratio of oropharyngeal cancer from the weakly informative prior and likelihood for each category of alcohol consumption. In these graphs, we overlay plots of the weakly informative prior, the likelihood function based on the observed data, and the posterior distribution based on Markov Chain Monte Carlo sampling. The observed data (summarized in the likelihood) are driving the estimation of the posterior distributions, and the weakly informative prior has modest impact on the

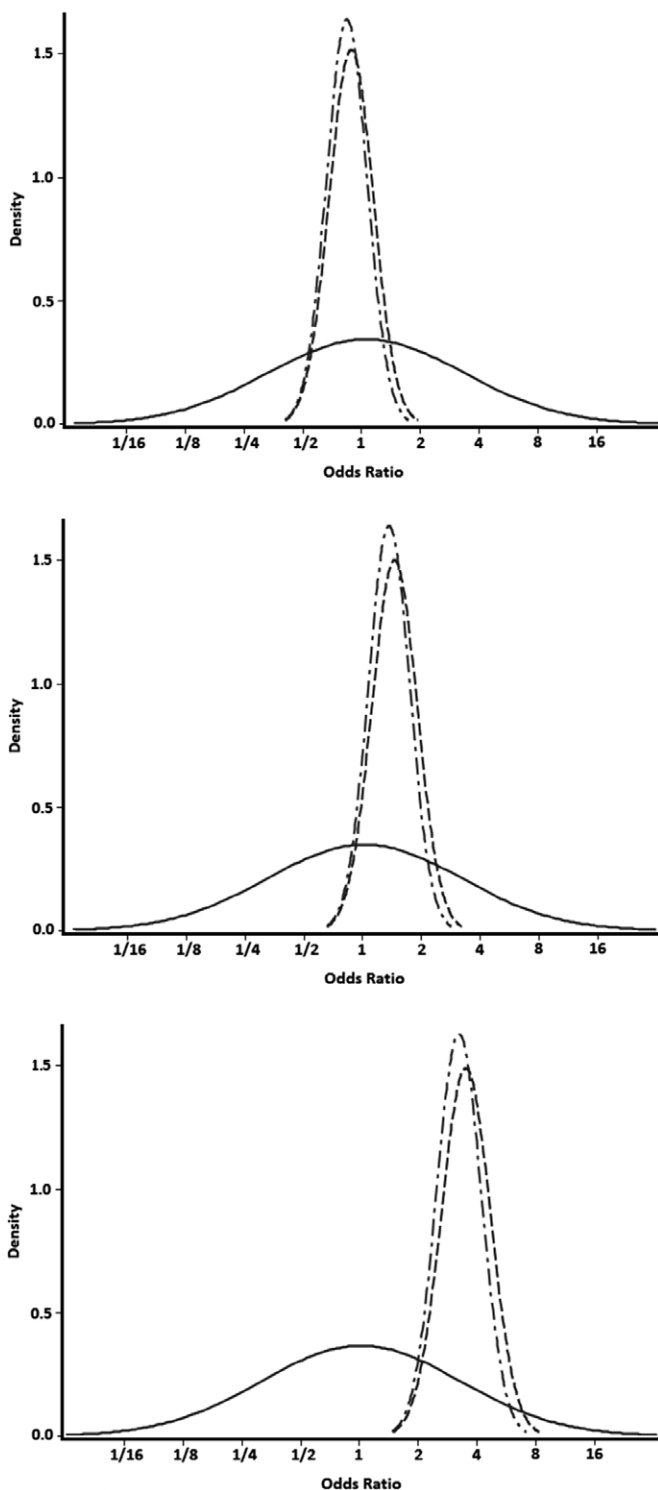


FIGURE 1. Kernel density plots for the weakly informative prior (solid line), likelihood (dashed line), and posterior (dash-dotted line) for the odds ratio of oropharyngeal cancer associated with categories of alcohol consumption represented by β_1 (upper), β_2 (middle), and β_3 (lower).

interpretation of the parameters. Only the highest tertile, where the magnitude and precision of the odds ratio change moderately, shows evidence of instability due to sparse data.

Unlike oropharyngeal cancer, the confidence intervals for the association of alcohol consumption and hypopharyngeal cancer are wide. The odds ratios associated with tertiles of lifetime alcohol consumption change by 105%, 114%, and 120%, respectively, when the weakly informative prior is implemented. In addition to a large shift in the parameter estimates toward the mean of the weakly informative prior, the upper bounds of the 95% PIs each show an approximately 10-fold decrease in magnitude, whereas the lower bound of the 95% PIs are largely unaffected. Figure 2 illustrates the relative contribution of the weakly informative prior and likelihood to the posterior distribution of the odds ratio of hypopharyngeal cancer. The likelihood and weakly informative prior provide a similar contribution to estimation of the posterior distributions, highlighting the extremely sparse nature of the available data.

The odds ratio for the lower tertile of alcohol consumption is shifted past the prior mean to a value of 0.79 (95% PI = 0.25, 2.61), which is a counterintuitive finding at first glance. This is a result of the extremely high correlation between parameters representing the effects of the highest and lowest tertiles of alcohol consumption (Pearson $r = 0.87$). The high correlation implies that if the odds ratio for the highest tertile is shrunk in one direction, the odds ratio for the lowest tertile will also be shrunk in that direction, even if the priors are independent. The substantial shrinkage of the highest tertile toward smaller values translated into additional shrinkage of the lowest tertile toward smaller values—in this case, values less than the null. Although the magnitude of these parameter estimates changes dramatically, decisions that might be based on statistical cutpoints represented by the 95% PI remain unchanged.

DISCUSSION

Quantitative techniques have been developed for post hoc evaluation of sensitivity of model parameters based on proposed degrees of confounding, misclassification, or selection bias.^{4,7,25} However, methods are undeveloped for quantifying the influence of adding modest information to the data, such as observing a few extra exposed and unexposed cases or controls. We have presented a simple approach for quantifying sensitivity of model results using a weakly informative prior based on general substantive beliefs about a credible range of values for the effect estimate of interest. Although the use of such priors has been proposed previously,^{26–28} limitations in statistical software have been a barrier to implementation. Advances in statistical software have now made appropriate tools easily accessible; one can incorporate weakly informative priors with the addition of a single line of software code (see the eAppendix, <http://links.lww.com/EDE/A650> for a simple example). Thus, this article also illustrates the relative ease with which analysis can include a Bayesian

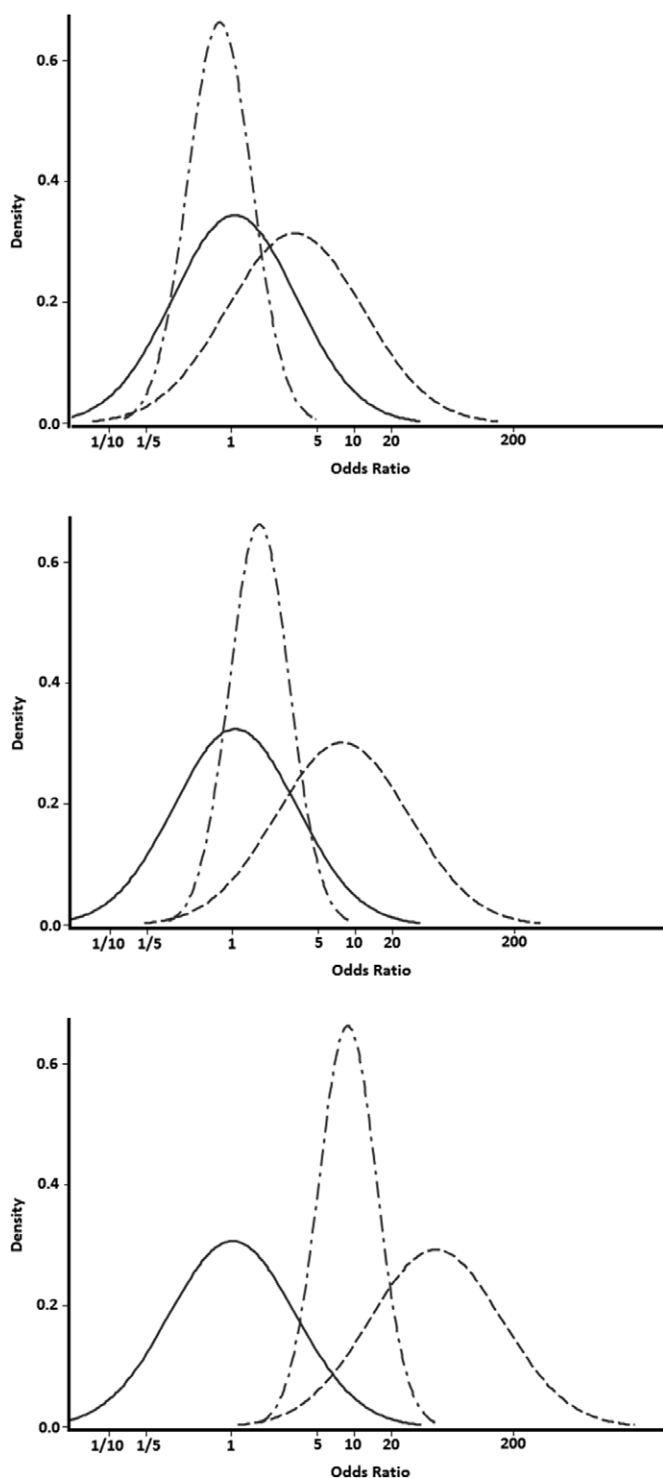


FIGURE 2. Kernel density plots for the weakly informative prior (solid line), likelihood (dashed line), and posterior (dash-dotted line) for the odds ratio of hypopharyngeal cancer associated with categories of alcohol consumption represented by β_1 (upper), β_2 (middle), and β_3 (lower).

component—a process that was previously quite difficult. Our examples include a case where the change in effect estimate is minimal, and another that tempers our interpretation of the magnitude of effect.

From a Bayesian perspective, frequentist regression models are often a special case of a Bayesian model—one in which a flat prior is specified and all values for a parameter estimate of interest are set a priori as equally plausible. However, most epidemiologists would not regard all values for parameters representing an exposure–disease relationship of interest as equally likely. Belief regarding a plausible range of values may be specific to a study of interest, based on the general body of knowledge in a substantive field (as in our example), or drawn from research in biology, toxicology, or even physics. It can be useful and important to recognize research external to our own, regardless of the source. Our weakly informative prior is an example of a simple way to quantitatively formalize this generic knowledge and to assess its impact on our results.

To minimize sparse-data problems when studying a specific exposure–disease relationship, researchers may attempt to simplify a regression model by systematically removing potential effect-measure modifiers or confounders.^{29,30} In some cases, this may be a reasonable approach. However, there are scenarios where model simplification may be untenable. Case-control studies often include matching designs to improve sampling efficiency, which conditions analyses on the matching factors.³⁰ Alternately, a researcher may believe specific variables and product terms need to be included in the regression models a priori based on substantive knowledge.^{31–33} In these cases, or when model reduction exercises fail to solve sparse-data problems, an approach to evaluate the sensitivity of a parameter estimate to sparse data can be valuable.

As with any analytic tool, a weakly informative prior faces limitations. First, although implementing a weakly informative prior with Markov Chain Monte Carlo can be easier than data augmentation, it requires familiarity with diagnosing Markov Chain Monte Carlo model convergence.³⁴ However, as Markov Chain Monte Carlo becomes more widely used, model convergence criteria will become better understood. Second, the strength of any informative prior, whether described as weak or strong, is inversely related to the weight of information provided by the data and specified regression model. As the data and model become more informative, the prior will become less informative. What may be viewed as weakly informative in some substantive settings may be viewed as overly informative or implausible in others. Therefore, it is important to consider specification of a weakly informative prior based on knowledge regarding the expectation of the magnitude, and possibly direction, of an etiologic relationship of interest.

Further, attention must be paid to the scale of variables in the model because a sensible weakly informative prior may

suddenly become nonsensical if the original variable is rescaled (eg, if it is divided by 100). As shown in the example, the use of a weakly informative prior (or any informative prior) on a single parameter can influence other parameters' estimates if there is high correlation between the parameter and the priors. In our example, the shift toward the prior mean for the effect of the highest tertile of alcohol consumption drives the estimate of the effect of the lowest category across the prior mean. Carlin and Louis³⁵ refer to this as "crossing" and describe its unpredictable occurrence as a consequence of integrating prior information into multivariable models. Although this is an unexpected result, it serves as a diagnostic check of the robustness of other parameters to modest changes to the information within the model. This type of result will often be an indication of sparse data as well as of high between-variable correlation. When crossing occurs, a research might consider using a weakly informative prior on individual parameters, rather than all model parameters.

The example was chosen because the sparseness of data is transparent. In many cases, sparse data may not be so obvious, particularly if it occurs in a confounder rather than the main exposure. Although good epidemiologic practice typically begins with univariate and bivariate descriptions of relevant variables, it may be impossible to examine all contingency tables in regression models that contain even moderate numbers of covariates. Further, what exactly constitutes sparseness is far from clear. Our weakly informative prior is designed to allow researchers to judge the impact of additional modest prior knowledge (or additional data) on their findings. Therefore, maximum likelihood estimates in the absence of a weakly informative prior should always be presented in addition to posterior estimates that use a weakly informative prior. This will also allow the use of the maximum likelihood estimates in future meta- or Bayesian analyses. When large-sample theory holds, the maximum likelihood estimate will be equal to a Bayesian estimate that uses a noninformative (or diffuse) prior for the parameters of interest. In standard epidemiologic regression models, such as logistic or log-binomial regression, sparse data can lead to estimates that are far from the truth. The use of informative priors, such as our proposed weakly informative prior, for correcting bias is well accepted by both frequentists and Bayesians as a way to potentially reduce mean squared error.^{36,37}

The use of a null-centered weakly informative prior is similar to ridge regression, which penalizes large parameter estimates in a regression model.³⁸ Other researchers have suggested a range of weakly informative priors based on different directions of magnitude, such as near the null, moderately positive, or moderately protective.^{19,27} In addition, Spiegelhalter et al³⁹ have advocated for a "skeptical" prior that weights the posterior parameter distribution toward a null effect (interpreted in a clinical setting as no difference between two treatments). Similar to the Cauchy prior recommended by Gelman et al,²⁶ we intend our prior to be broadly applicable

by epidemiologists. If the desire is to inform parameter estimation with a narrower or broader range of parameter values, an analyst can simply adapt the specified variance. Further, it is possible to specify a weakly informative prior for some parameters in the data and not others. Specifying a range of weakly informative priors will increase the researcher's understanding of a parameter's sensitivity to the addition of different information, whether it is more or less precise or centered on a protective or harmful estimate of the effect. We suggest a range with 95% of the prior mass of relative values between 0.1 and 10 as a starting point. However, when more (or less) informative priors are supported by evidence in the existing literature, it would be recommended to apply them in addition, or as an alternative, to the weakly informative prior specified here.

In our example, a reasonable conclusion would be that the parameter estimates are too unstable for reliable decision making or inference. If parameter estimates are unchanged with a weakly informative prior, one might conclude that the results of the original analysis are robust to additional, external information and thus more useful to policy makers. As with other sensitivity analyses, the ultimate benefit of a weakly informative prior is to provide a better understanding of the strengths and limitations of the data on which decisions or inference are based.

ACKNOWLEDGMENTS

We thank Andrew Olshan and Anne Hakenewerth for providing an example and comments on this article.

REFERENCES

- Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol*. 2004;160:301–305.
- MacLehose RF, Dunson DB, Herring AH, Hoppin JA. Bayesian methods for highly correlated exposure data. *Epidemiology*. 2007;18:199–207.
- Greenland S. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics*. 2000;1:113–122.
- Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational data. *Epidemiology*. 2003;14:451–458.
- Gustafson P, Le ND, Saskin R. Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics*. 2001;57:598–609.
- Chu R, Gustafson P, Le N. Bayesian adjustment for exposure misclassification in case-control studies. *Stat Med*. 2010;29:994–1003.
- 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;160:384–392.
- Dunson DB. Commentary: practical advantages of Bayesian analysis of epidemiologic data. *Am J Epidemiol*. 2001;153:1222–1226.
- Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med*. 1999;130:1005–1013.
- Wacholder S, Samet J. Consequences of "big epidemiology". *Am J Epidemiol*. 2006;163:S169–S169.
- Breslow N. Odds ratio estimators when the data are sparse. *Biometrika*. 1981;68:73–84.
- Breslow NE, Day NE. Statistical methods in cancer research. Volume 1 - The analysis of case-control studies. *IARC Sci Publ*. 1980;32:5–338.
- Greenland S, Schwartzbaum JA, Finkle WD. Problems due to small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol*. 2000;151:531–539.

14. Greenland S, Christensen R. Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression. *Stat Med*. 2001;20:2421–2428.
15. Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *Int J Epidemiol*. 2007;36:195–202.
16. Greenland S. Simpson's paradox from adding constants in contingency tables as an example of Bayesian noncollapsibility. *Am Stat*. 2010;64:340–344.
17. Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet*. 1955;19:251–253.
18. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch Environ Health*. 1994;49:9–16.
19. Witte JS, Greenland S, Kim LL. Software for hierarchical modeling of epidemiologic data. *Epidemiology*. 1998;9:563–566.
20. Hakenewerth AM, Millikan RC, Rusyn I, et al. Joint effects of alcohol consumption and polymorphisms in alcohol and oxidative stress metabolism genes on risk of head and neck cancer. *Cancer Epidemiol Biomarkers Prev*. 2011;20:2438–2449.
21. Rothman K, Keller A. The effect of joint exposure to alcohol and tobacco on risk of cancer of the mouth and pharynx. *J Chronic Dis*. 1972;25:711–716.
22. Keller AZ, Terris M. The association of alcohol and tobacco with cancer of the mouth and pharynx. *Am J Public Health Nations Health*. 1965;55:1578–1585.
23. Arminger G, Clogg CC, Sobel ME. *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum Press; 1995.
24. Hoerl AE, Kennard RW. Ridge regression—Applications to nonorthogonal problems. *Technometrics*. 1970;12:69–82.
25. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. *Ann Epidemiol*. 2006;16:834–841.
26. Gelman A, Jakulin A, Pittau MG, Su YS. A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat*. 2008;2:1360–1383.
27. Greenland S. Putting background information about relative risks into conjugate prior distributions. *Biometrics*. 2001;57:663–670.
28. King G, Zeng L. Logistic regression in rare events data. *Political Analysis*. 2001;9:137–163.
29. Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health*. 1989;79:340–349.
30. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
31. Weng HY, Hsueh YH, Messam LL, Hertz-Picciotto I. Methods of covariate selection: directed acyclic graphs and the change-in-estimate procedure. *Am J Epidemiol*. 2009;169:1182–1190.
32. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37–48.
33. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002;155:176–184.
34. Cole SR, Chu H, Greenland S, Hamra G, Richardson DB. Bayesian posterior distributions without Markov chains. *Am J Epidemiol*. 2012;175:368–375.
35. Carlin BP, Louis TA. *Bayesian Methods for Data Analysis*. 3rd ed. Boca Raton: CRC Press; 2009.
36. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 2000;42:80–86.
37. Greenland S. Principles of multilevel modelling. *Int J Epidemiol*. 2000;29:158–167.
38. Hoerl AE. Ridge regression. *Biometrics*. 1970;26:603.
39. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health Care Evaluation*. Chichester, Hoboken, NJ: Wiley; 2004.