Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2013;42:308–317
doi:10.1093/ije/dys213


# Bayesian regression in SAS software

**Sheena G Sullivan[1,2]* and Sander Greenland[1,3]**

[1]Department of Epidemiology, University of California, Los Angeles, CA, USA, [2]WHO Collaborating Centre for Reference and Research on Influenza, Melbourne, Australia and [3]Department of Statistics, University of California, Los Angeles, CA, USA

*Corresponding author. WHO Collaborating Centre for Reference and Research on Influenza, 10 Wreckyn St, North Melbourne, Victoria, Australia. E-mail: sgsullivan@ucla.edu/sheena.sullivan@influenzacentre.org

**Accepted** 13 November 2012

Bayesian methods have been found to have clear utility in epidemiologic analyses involving sparse-data bias or considerable background information. Easily implemented methods for conducting Bayesian analyses by data augmentation have been previously described but remain in scant use. Thus, we provide guidance on how to do these analyses with ordinary regression software. We describe in detail and provide code for the implementation of data augmentation for Bayesian and semi-Bayes regression in SAS® software, and illustrate their use in a real logistic-regression analysis. For comparison, the same model was fitted using the Markov-chain Monte Carlo (MCMC) procedure. The two methods required a similar number of steps and yielded similar results, although for the main example, data augmentation ran in about 0.5% of the time required for MCMC. We also provide online appendices with details and examples for conditional logistic, Poisson and Cox proportional-hazards regression.

**Keywords** Bayesian methods, data augmentation, Markov-chain Monte Carlo (MCMC), logistic models

## Introduction

The vast majority of published statistical analyses are frequentist: data probabilities are calculated from models, and these probabilities are used to make inferences about model parameters. Bayesian methods, on the other hand, combine those data models with prior probability distributions for model parameters, to produce posterior probabilities for those parameters. The need to specify a prior probability distribution has led to criticism that Bayesian methods are too subjective, with claims that frequentist methods provide an objective evaluation of the data. In nonexperimental research, however, frequentist data models are of doubtful validity, so that the 'objectivity' of frequentist methods means only that they have become the accepted convention; they do not encompass any absolute scientific standard of objectivity or validity.[1–5]

With careful choice of prior distributions, Bayesian methods can outperform traditional frequentist methods under frequentist criteria of mean-squared error of estimates and interval-estimate coverage, especially when the data are sparse (e.g. when the number of covariates approaches the numbers of exposed cases or noncases) or biases are present that cannot be handled by conventional methods.

Despite their advantages, Bayesian methods remain uncommon in many disciplines. This is likely due in part to their absence from most basic statistics curricula, and because until recently major statistical packages did not have explicit procedures for performing Bayesian analyses. Now, however, SAS® software (versions 9.2 onward) provides a Markov-chain Monte-Carlo (MCMC) procedure. A simpler option is data augmentation,[6–11] in which each prior distribution is represented by one or more prior-data records. This method allows one to conduct Bayesian analyses with any regression software, thus eliminating the need to purchase or rely on unfamiliar programs. It does not require use of priors on all coefficients, so can be used to conduct semi-Bayes (partial-Bayes) analyses. Data augmentation also runs much faster than simulation methods like MCMC, and (unlike MCMC) it introduces no complex convergence

concerns, yet in regression examples to date it has given similar results.[6–8]

Although applications have begun to appear outside methods articles,[12] adoption has been slow, possibly because clear guidance on how to do so is lacking. Thus, we describe the application of Bayesian regression in SAS software using a well-studied clinical cohort.[7,9] We compare the results obtained from data augmentation with those derived from MCMC, giving detailed code in the Appendix. Online appendices give further examples including data and full code for Bayesian conditional logistic, Poisson log-linear and Cox regression (Supplementary data are available at *IJE* online).

All Bayesian methods require the extra labour of specifying prior distributions, and entail a special hazard in that using a poorly chosen prior distribution may degrade the performance of inferential procedures. Nonetheless, it has long been observed that neutral, weakly informative prior distributions lead to better frequentist performance in the kind of applications typical in epidemiology, especially when the data are sparse.[13–16] Sparse-data artefacts often go unrecognized in study reports; for example, one case-control analysis[17] gave an odds ratio of 65 (95% CL 6.3, 672) for the association of ever smoked with ICU admission in a logistic regression involving 12 regressors and 120 cases, with no comment on the implausible nature of the estimate. When the estimate is instead plausible, bias will go completely unnoticed without further analysis. Thus, we will also describe how weakly informative priors can be easily created to diminish sparse-data artefacts without requiring excessive contextual information.

## Data augmentation via the offset method in SAS software: a tutorial with informative priors

The example data came from a study of obstetric care and neonatal death at a teaching hospital.[18] There were 14 covariates (regressors) available for study. Death is uncommon enough in all subgroups so that odds ratios approximate risk ratios. We thus use logistic regression, so that the coefficient $\beta$ of a covariate X represents the log odds ratio, $\ln(OR)$, or the change in log odds per unit change in X, with the odds ratio $OR$ equal to $e^{\beta}$.

For these data, a logistic model (PROC LOGISTIC) with all 14 variables converges without problem, but produces some extremely inflated estimates due to sparse-data bias (Table 1). The most extreme example of inflation is the indicator of hydramnios during pregnancy, with a maximum-likelihood (ML) estimate for the odds ratio of $OR = 60$, 95% Wald confidence limits 5.7, 635 and 95% profile-likelihood limits 2.8, 478, reflecting only one death among nine hydramnios pregnancies. Hence it makes sense to shrink the estimates back to something more reasonable, for example by imposing penalty functions or priors on the coefficients to be shrunk.

We describe this process for a class of priors for the log odds ratio $\beta$, a class that includes normal priors for $\beta$ and hence lognormal priors for $OR = e^{\beta}$ (data representing asymmetric priors can also be created[8,10,19]). These priors are symmetric with one mode (peak), hence the mean, median and mode equal the same value, which we call the prior centre and denote by $\beta_{prior}$; its antilog $OR_{prior} = \exp(\beta_{prior})$ is the prior median odds ratio (the prior on the odds-ratio scale is skewed rather than symmetric, so that $OR_{prior}$ is neither the prior mode nor prior mean of the odds ratio). The prior-data record for $\beta$ is constructed so that if $\beta$ were estimated by fitting the no-intercept logistic model $\Pr(Y = 1|X = x) = e^{\beta x}/(1 + e^{\beta x})$ to this record alone, its point and variance estimates would be the centre $\beta_{prior}$ and variance $v_{prior}$ of the prior distribution for $\beta$.

Upon using a logistic-regression program to analyse the augmented data set (which contains both the actual and prior data records) the 'ML estimate' for $\beta$ from the program output is the posterior mode $\beta_{post}$ of $\beta$; $\beta_{post}$ is also an approximate posterior median and mean for $\beta$. The 'standard error' from the output is the approximate posterior standard deviation for $\beta$, $s_{post}$. The 'odds ratio estimate' and its '95% confidence limits' from the output are the approximate posterior median $OR_{post} = \exp(\beta_{post})$ and 95% posterior limits (2.5th and 97.5th posterior percentiles). We show results from two types of interval approximations available in SAS: the default Wald interval $\exp(\beta_{post} \mp 1.96 s_{post})$, and the more accurate profile-likelihood interval provided by the <clodds = both> option in the model statement. Using approximately normal priors, the accuracy of these approximations to posterior intervals is at least as good and often much better than the coverage accuracy of the corresponding likelihood intervals from the unaugmented data.

### Constructing the prior data

Creation of a normal prior for $\beta$ requires specifying a prior interval ($OR_{lower}$, $OR_{upper}$) for its corresponding odds ratio $OR = e^{\beta}$, within which we would bet a given percent (usually 95%) that $\beta$ lies inside. The prior centre $\beta_{prior}$ is then just the average of the log limits:

$$\beta_{prior} = \ln(OR_{prior}) = \frac{\ln(OR_{upper}) + \ln(OR_{lower})}{2}$$

The prior variance $v_{prior}$ of $\beta$ is the squared width of the log interval expressed in standard-deviation units, which for a normal 95% interval is:[4]

$$v_{prior} = \left[\frac{\ln(OR_{upper}) - \ln(OR_{lower})}{2 \times 1.96}\right]^2$$

To illustrate, one plausible prior for the hydramnios risk ratio would assign 2:1 odds (67% probability) to

**Table 1** Multiple logistic regressions of neonatal-death risk in a cohort of 2992 births with 17 deaths, intercept and 14 regressors in each model. Shown are the prior median odds ratio $OR_{prior}$ and 95% limits; ML estimates with 95% Wald and profile-likelihood (profile) limits; approximate posterior medians from data augmentation including a prior on all 14 regressors with 95% Wald and profile limits, using prior data with ½ correction (A = 4.5) or with rescaled prior data (S = 10, A = 400); and simulated posterior medians and 95% limits (2.5th and 97.5th percentiles) from MCMC with normal priors

| | | | | Approximate posterior median (95% posterior limits) | | |
|---|---|---|---|---|---|---|
| **Regressor ($X_j$)** | **Deaths with $X_j > 0$** | **Prior median $OR_{prior}$ (95% prior limits)** | **ML estimate: A = 0 (95% Wald and profile limits)** | Data augmentation: A = 4.5[a] | Data augmentation with a rescaled ($S = 10$) prior[a] | MCMC[b] |
| Non-White | 5 | 2 (0.5,8) | 1.9 (0.55,6.5) (0.51,6.3) | 1.8 (0.75,4.2) (0.72,4.1) | 1.8 (0.73,4.3) (0.71,4.2) | 1.8 (0.70,4.2) |
| Early age | 3 | 2 (0.5,8) | 1.6 (0.39,6.7) (0.32,6.1) | 1.6 (0.65,4.1) (0.62,3.9) | 1.6 (0.63,4.1) (0.61,4.0) | 1.6 (0.59,4.0) |
| Nulliparity | 8 | 2 (0.5,8) | 1.5 (0.51,4.7) (0.50,4.9) | 1.6 (0.69,3.5) (0.68,3.6) | 1.5 (0.67,3.6) (0.67,3.6) | 1.6 (0.67,3.6) |
| Gestational age | 10 | 4 (1,16) | 4.9 (2.4,9.8) (2.4,10.0) | 4.5 (2.5,8.0) (2.5,8.1) | 4.5 (2.5,8.1) (2.5,8.1) | 4.6 (2.5,8.3) |
| Isoimmunization | 1 | 2 (0.5,8) | 3.0 (0.91,10) (0.62,8.5) | 2.4 (0.95,6.0) (0.87,5.6) | 2.4 (0.94,6.2) (0.85,5.7) | 2.3 (0.81,5.6) |
| Past abortion | 2 | 1 (0.25,4) | 0.72 (0.18,2.9) (0.12,2.3) | 0.84 (0.34,2.1) (0.31,1.9) | 0.83 (0.33,2.1) (0.31,1.9) | 0.79 (0.29,1.9) |
| Hydramnios | 1 | 4 (1,16) | 60 (5.7,635) (2.8,478) | 5.8 (1.6,21) (1.6,22) | 6.1 (1.6,23) (1.6,23) | 6.0 (1.6,22) |
| Labour progress | 2 | 2 (0.5,8) | 0.50 (0.06,3.9) (0.04,2.8) | 1.3 (0.45,3.5) (0.42,3.3) | 1.2 (0.43,3.5) (0.41,3.3) | 1.2 (0.40,3.3) |
| PCA | 1 | 2 (0.5,8) | 3.1 (0.33,29) (0.15,20) | 2.2 (0.71,7.1) (0.67,7.0) | 2.3 (0.68,7.5) (0.65,7.2) | 2.2 (0.64,7.1) |
| No monitor | 3 | 2 (0.5,8) | 1.2 (0.32,4.9) (0.35,5.9) | 1.8 (0.68,4.5) (0.70,4.7) | 1.7 (0.66,4.6) (0.68,4.8) | 1.8 (0.71,5.0) |
| Twin, triplet | 3 | 4 (1,16) | 8.2 (1.8,37) (1.5,33) | 5.1 (1.9,14) (1.8,14) | 5.2 (1.9,15) (1.8,14) | 5.3 (1.8,14) |
| Public ward | 6 | 2 (0.5,8) | 0.86 (0.26,2.9) (0.25,2.8) | 1.3 (0.56,3.0) (0.54,3.0) | 1.3 (0.54,3.0) (0.53,3.0) | 1.3 (0.53,3.0) |
| PROM | 1 | 2 (0.5,8) | 0.54 (0.06,4.8) (0.03,3.2) | 1.3 (0.45,3.5) (0.41,3.3) | 1.2 (0.43,3.5) (0.41,3.3) | 1.2 (0.39,3.3) |
| Malpresented | 3 | 4 (1,16) | 3.9 (0.88,17) (0.73,15) | 3.9 (1.5,10.0) (1.4,9.8) | 3.9 (1.4,10) (1.4,9.9) | 3.8 (1.4,10.0) |

ML: Maximum likelihood; MCMC: Markov-chain Monte Carlo; PCA: placental or cord anomaly. PROM: premature rupture of membranes.
Variables are indicators except early age (0 = 20+, 1 = 15–19, 2 = under 15), gestational age (0 = no, 1 = 36–38 weeks, 2 = 33–36 weeks; under 33 weeks excluded), isoimmunization (0 = no, 1 = Rh, 2 = ABO), labour progress (0 = no, 0.33 = prolonged, 0.67 = protracted, 1 = arrested) and past abortion (0 = none, 1 = 1, 2 = 2+).
[a]Limits shown are Wald exp(estimate $\mp$ 1.96 × standard error) and profile likelihood from PROC LOGISTIC.
[b]Number of MCMC samples was 100 000

this ratio falling between 2 and 8, and 95% probability to this ratio falling between 1 and 16. This leads to a prior centre of $\beta_{prior} = \ln(OR_{prior}) = \frac{\ln(16)+\ln(1)}{2} = \ln(4)$ and a prior variance of $v_{prior} = \left[\frac{\ln(16)-\ln(1)}{2\times 1.96}\right]^2 = 0.5$. This prior reflects the belief that hydramnios could cause neonatal death, with a risk ratio (and hence odds ratio) probably in the range of 1 to 16.

We need to go through this prior assignment exercise only for variables for which we want to use a prior. Depending on the original data format and the chosen priors, one or more new variables may also be needed. Specifically, the final data set will need the following variables:

(1) A = the number of cases represented by the data record; in the foetal monitoring data, A is the number of deaths. For individual (ungrouped) data, A is 1 for all actual cases and 0 for all actual noncases. To determine A for a record representing a normal prior distribution with variance $v_{prior}$, note that the approximate variance a logistic program associates with the record is 1/A+1/A = 2/A, which will equal $v_{prior}$ upon setting A = $2/v_{prior}$. In PROC LOGISTIC, this variable A is the 'events' in the 'events/trials' model statement:[20]

MODEL events/trials =< effects >

(2) M = the total number of subjects represented by the record. For ungrouped data, M is set to 1 for all actual-data records. For a prior record, M = 2A. Thus the proportion of cases in the record is A/M, which in ungrouped actual-data records is 1/1 = 1 for cases and 0/1 = 0 for noncases, and is A/2A = ½ in the prior records. In PROC LOGISTIC, this variable M is the 'trials' in the 'events/trials' model statement.

(3) Using the noint option, the program is requested not to force in its own intercept. An indicator variable *Const* replaces the constant: *Const* is set to 1 for all actual-data records and for the prior record for the intercept (if such a prior is included); *Const* is set to 0 for all other prior records. *Const* is then included as a regressor in the model. Its coefficient is the intercept (usually denoted $\alpha$ or $\beta_0$), which is the log odds of being a case when all other regressors are 0; it is thus important to make sure that 0 is a meaningful value for all regressors.

(4) If the prior centre $\beta_{prior}$ is not zero for some regressors (so some shrinkage is not toward zero), we will also need to add an 'offset' variable $H$ (an offset is a variable that is entered directly into the model with its coefficient set equal to 1 by the program). $H$ is specified as an offset in the model step by using the offset=H option. To define $H$, set $H = 0$ for all actual records (unless the actual data contain offsets, in which case $H$ equals those offsets); for prior-data records, set $H = -\beta_{prior}$. If, however, all the priors are centred at zero, $H$ will be zero everywhere and need not be added to the data.

(5) For the actual data, no change is made to the regressor (covariate) values. For the prior-data record for coefficient $\beta$ of regressor X, we set X = 1 to indicate that the record represents the prior for $\beta$; all other regressors in this prior record are set to 0. As a result, all the prior-data records have regressor entries that are zero everywhere except for one entry of 1 that identifies the regressor to which the record refers.

Use of no prior information for a coefficient corresponds to infinite $v_{prior}$ and thus A = 0, M = 0. Such a record would not be used for the model fitting, which is why we can omit prior records for coefficients to which we assign no prior. Examples of informative log-normal priors are given in Table 2. The prior implied by the data record becomes more normal as A increases, being practically normal for A ⩾ 5 and having notably heavier tails (more spread) than normal for A < 4.

For the hydramnios coefficient, using a rough normal approximation for the log odds ratio,[9,21] a $v_{prior} = 0.5$ produces a prior-data record containing A = $2/v_{prior} = 2/0.5 = 4$ deaths and M = 8 subjects total. A prior median for $OR_{prior}$ of 4 translates to a prior centre of $\beta_{prior} = \ln(4)$ and an offset of H = $-\beta_{prior} = -\ln(4) = -1.39$. The record also has hydram = 1 (which indicates the prior record is for the hydramnios coefficient); all other variables in the record (including *Const*) are set to 0 (see Appendix and Table 2).

## Improving the prior-interval approximation

Paralleling analysis of actual tabular data,[13,14] the approximation in step 1 can be improved slightly by adding ½ to each prior count A, making A = $2/v_{prior} + ½$. For example, the exact prior probability that $1 < OR < 16$ implied by a record with A = 4, M = 8, is only 93%; adding 0.5 to A we get A = 4.5 and M = 9, for an exact prior probability of 95% for $1 < OR < 16$. We use this improvement in the present analysis. For user convenience, more precise conversions from 95% prior limits for an odds ratio to the prior case-count A are as follows: (1/40,40) becomes A = 1; (1/16,16) becomes A = 1.5; (1/8,8) becomes A = 2.3; (1/5,5) becomes A = 3.5; (1/4,4) becomes A = 4.5; (1/3,3) becomes A = 6.9; (1/2,2) becomes A = 16.6; (0.67,1.5) becomes A = 47; (0.80,1.25) becomes A = 155; and (0.83,1.2) becomes A = 232. These figures are derived by noting that a prior-data record with M = 2A induces an F(M,M) prior distribution for $e^{\beta}$.[10,19] We caution however that some programs may truncate input counts and so remove the fraction after the decimal point; thus one should have the program print back the input counts to check if the program is truncating fractions.

**Table 2** Examples of log-normal priors with corresponding records for augmentation for unconditional logistic regression

| Desired prior for ln(*OR*)[a] | Prior median $OR_{prior}$ | Prior variance $v_{prior}$ of ln(*OR*) | 95% limits | | *S* | Cases (deaths) A | Total M = 2A | Covariate value[b] | Offset (H) -ln($OR_{prior}$) |
|---|---|---|---|---|---|---|---|---|---|
| | | | Lower | Upper | | | | | |
| N(ln(2),0.5) | 2 | 0.5 | 0.5 | 8 | 1 | 4.5 | 9 | 1 | −0.693 |
| | | | | | 10 | $4 \times S^2 = 400$ | 800 | $1/S = 0.1$ | $-0.693/S = -0.0693$ |
| N(ln(4),0.5) | 4 | 0.5 | 1 | 16 | 1 | 4.5 | 9 | 1 | −1.39 |
| | | | | | 10 | 400 | 800 | 0.1 | −0.139 |
| N(0,0.5) | 1 | 0.5 | 0.25 | 4 | 1 | 4.5 | 9 | 1 | 0 |
| | | | | | 10 | 400 | 800 | 0.1 | 0 |

[a]N($\mu$,v) = normal prior with mean $\mu$, variance v. The three priors listed are those used for the main effects in the example. For each prior, the second row illustrates the record entries when using a scale factor $S = 10$ to improve the normal approximation.
[b]Value for covariate to which prior refers. All other covariates are zero.

In our analyses, priors were placed on all model coefficients except the intercept (this differs from Greenland,[7,9] where an intercept prior was also used) (Table 1). Greenland[7] describes the background rationale for the chosen priors. The fifth column in Table 1 shows the results from running the regression using these prior data in the SAS software code in the Appendix. Using the model with *Const* and all 14 regressors, the approximate posterior median and limits obtained from PROC LOGISTIC for the hydramnios coefficient were $OR_{post} = 5.8$, Wald limits 1.6, 21, profile limits 1.6, 22, all much more reasonable than the ML results. Profile limits are preferable and can be obtained using the <clodds = pl> option in the model statement for the logistic procedure, or <lrci> in PROC GENMOD.

### Rescaling the prior

Although the addition of ½ to the prior count A brings the prior coverage very close to the stated value and is adequate for all but the smallest A, perfectly normal priors can be imposed by using a rescaling factor *S* that is divided into all the regressor values in the prior data, including the offset *H*; the prior A and M are then inflated by a factor of $S^2$ to compensate.[7,9] Thus, using $S = 10$, the prior record for hydramnios has $A = 4 \times S^2 = 400$, $M = 800$, $Const = 0$, $H = -1.39/S = 0.139$, and $hydram = 1/S = 0.1$. As can be seen in Table 1 (sixth column), this rescaling approach made little difference in this example: the results for the hydramnios coefficient were $OR_{post} = 6.1$, Wald limits 1.6, 23, profile limits 1.6, 23, and we would not expect it to make much practical difference compared with adding ½ to A unless the initial $A = 2/v_{prior}$ was under 3. We caution however that to use this approach with a polytomous variable (an unordered categorical variable) one should not declare the variable as categorical to the program, but should instead create the binary category indicators for the variable as new (dummy) variables and use those for the regression, discarding the original variable. Otherwise, the program will treat the value $1/S$ as defining a new category for the variable and will thus not impose the desired prior.

## Comparison with Markov-chain Monte-Carlo via the SAS/STAT software Bayes statement

Since version 9.2, SAS/STAT software has been capable of MCMC Bayesian analyses through the addition of a BAYES statement in the GENMOD, PHREG and LIFEREG procedures, and has a general purpose MCMC procedure. We re-ran the above analysis using the BAYES statement in GENMOD (see Appendix for syntax) and obtained results similar to data augmentation (Table 1, last column). For the hydramnios coefficient, the MCMC posterior median for the odds ratio was $OR_{post} = 6.0$ with 95% limits 1.5, 22, exceptionally good agreement considering there is simulation error in MCMC and approximation error in data augmentation.

Like data augmentation, MCMC requires labour beyond standard programming. For MCMC, SAS/STAT software requires the user to create a data set containing the prior means and variances for each coefficient. If none is desired, the documentation suggests using a noninformative N(0,$10^6$) prior for the log odds ratio. This differs from augmentation where one can simply choose to not include a prior record for a particular coefficient. Additionally, MCMC demands much more run time. On a 64-bit machine with 2 Gb RAM with Windows 7, GENMOD fitted the model containing all 14 main effects with profile limits in about 30 s using data augmentation, but required more than 1 h using MCMC via the BAYES statement just to reach nominal convergence and generate enough samples to set limits. Another posterior simulation method, importance sampling, does not have convergence issues but is computationally more complex and intensive than data augmentation.[6]

## Weak priors for estimate stabilization

Suppose one wishes to avoid the labour or possible controversy associated with informative priors such as those used above. Weak 'smoothing' priors can still be justified as frequentist devices to stabilize estimates, reducing both their sparse-data bias and variance and thus increasing their accuracy.[13–16] These benefits of using priors also facilitate the inclusion of more confounder terms in the model, thus allowing further potential for bias reduction.[22] One simple approach is to modify the contingency-table tradition of adding a constant to each cell (which can lead to paradoxical effects on estimates[23]) by instead using prior-data records derived from very diffuse coefficient priors centred around 0 so that no offset is needed. This small addition to the data removes the most extreme sparsity, thus reducing bias (Table 3).

In the 18th century, Laplace (reported in Good[2]) proposed a 'law of succession' for estimating the chance of success in binomial trials by adding 1 to each of the numbers of observed successes and failures, which corresponds to a uniform prior on the chance of success. This approach generalizes to estimating logistic-regression coefficients by adding a prior record with $A = 1$, $M = 2$ for the coefficient.[22,23] This record corresponds to a logistic prior distribution that places about 95% probability on the odds ratio $e^{\beta}$ falling between 1/40 and 40, and thus is very weak.

The final two columns of Table 3 show the results of applying this prior to the example data, both when zero-centred and non-zero-centred priors are used. As expected, for fairly stable coefficients like gestational age the choice among these moderate to weak priors makes no practical difference in the final estimates, whereas for very unstable coefficients like hydramnios the choice of prior centre makes a large difference. The sensitivity analysis thus warns us that any inference of 'significance' in these unstable cases is highly dependent on choice of prior, with no prior (the ML result) being potentially as or more misleading than any reasonable, modestly informative prior. For comparison, a non-Bayesian sparse-data adjustment is available in SAS using the FIRTH command in the logistic procedure. Using this command, the odds ratio for hydramnios became 68 with 95% confidence limits of 6.1, 421, which is still unsatisfactory; hence we prefer the prior-data approach.

## Discussion

Data augmentation is a form of penalized-likelihood estimation (PLE) in which the prior data forces the program to generate a penalty function which imposes the prior constraints. Macros have been written for PLE in SAS software (e.g. FL, CFL and FLPM),[24] but are less straightforward than either of the two methods described here and require the SAS matrix language, SAS/IML. Data augmentation offers some advantages over both PLE macros and MCMC, not only in speed but also for interpretation: the initial $A = 2/v_{prior}$ shows the strength of the prior being imposed in terms of number of subjects added.[9,21] It is also readily transferable between different software platforms and needs only standard procedures taught in statistics classes, such as logistic regression. The syntax in the Appendix shows how these prior data records are appended to the actual data; this can be done with any logistic regression procedure which allows grouped data, specification of an offset, and suppression of the intercept. If no 'offset' option is provided by the program, one can impose the option by adding the offset variable and using one more prior-data record to constrain its coefficient to equal 1.[9] If one wants a skewed prior, this is easily accomplished by using an M unequal to 2A (although with skewed priors, use of profile-likelihood limits will be essential).[9,10] Dependent priors can also be transformed into prior data.[8,9] Finally, as illustrated by examples in the online Appendix (Supplementary data are available at *IJE* online), data augmentation can also be easily applied to conditional logistic, Poisson and Cox proportional hazards regression.

It is possible to include latent variables in the model by adding them to the data set with missing-data codes, along with their prior records, and fitting the model using a missing-data program (as in bias analysis).[19] In these applications there is usually little or no identification of key parameters under realistic models. As a result, ML will break down completely, and the approximations underpinning point estimates and Wald limits from PLE and hence data augmentation may also break down. Furthermore, both PLE and MCMC may suffer convergence problems; hence good convergence diagnostics will be essential.

Although not an issue in our example, in general to facilitate interpretation of coefficients and their priors it will be important to have all quantitative variables recentred to ensure that zero is a meaningful reference value present in the data, and rescaled so that their units are meaningful differences spanning a range present in the data. These transformations can also reduce the risk of convergence problems. For example, diastolic blood pressure could be recentred so that 0 represents 80 mm, and then rescaled to cm instead of mm, so that 95 mm would become $(95 - 80)/10 = 1.5$; adult age could be recentred so that 0 represents age 60 years, and rescaled to decades instead of years, so that 80 years would become $(80 - 60)/10 = 2.0$.

We again caution that for the intercept to be meaningfully interpreted, all the regressors must have zero as a meaningful value. Furthermore, regardless of Bayesian method, it may often be preferable to omit the intercept prior, not only when prior information on the population intercept is lacking, but also when

**Table 3** Prior-sensitivity analysis of the neonatal death data to illustrate the influence of the prior on the posterior estimates

| Regressor ($X_j$) | Deaths with Xj>0 | Prior 1: A = 4.5 prior median $OR_{prior}$ (95% prior limits) | Prior 2: A = 1 prior median $OR_{prior}$ (95% prior limits) | ML estimate: A = 0 (95% Wald and profile-likelihood limits) | Approximate posterior median (95% Wald limits) (95% profile limits) for odds ratio from data augmentation | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | A = 4.5, nonzero-centred coefficient prior | A = 4.5, zero-centred coefficient prior | A = 1, nonzero-centred coefficient prior | A = 1, zero-centred coefficient prior |
| Non-White | 5 | 2 (0.5,8) | 2 (0.13,32) | 1.9 (0.55,6.5) (0.51,6.3) | 1.8 (0.75,4.2) (0.72,4.1) | 1.3 (0.54,3.1) (0.53,3.1) | 1.8 (0.60,5.4) (0.57,5.3) | 1.6 (0.53,4.8) (0.50,4.7) |
| Early age | 3 | 2 (0.5,8) | 2 (0.13,32) | 1.6 (0.39,6.7) (0.32,6.1) | 1.6 (0.65,4.1) (0.62,3.9) | 1.2 (0.45,3.0) (0.43,3.0) | 1.6 (0.46,5.3) (0.40,5.0) | 1.3 (0.38,4.7) (0.34,4.4) |
| Nulliparity | 8 | 2 (0.5,8) | 2 (0.13,32) | 1.5 (0.51,4.7) (0.50,4.9) | 1.6 (0.69,3.5) (0.68,3.6) | 1.2 (0.54,2.6) (0.53,2.7) | 1.5 (0.55,4.1) (0.54,4.2) | 1.4 (0.51,3.7) (0.50,3.7) |
| Gestational age | 10 | 4 (1,16) | 4 (0.25,64) | 4.9 (2.4,9.8) (2.4,10.0) | 4.5 (2.5,8.0)(2.5,8.1) | 4.0 (2.2,7.3) (2.2,7.3) | 4.7 (2.4,8.9) (2.4,9.0) | 4.6 (2.4,8.8) (2.4,8.9) |
| Isoimmunization | 1 | 2 (0.5,8) | 2 (0.13,32) | 3.0 (0.91,10) (0.62,8.5) | 2.4 (0.95 ,6.0) (0.87,5.6) | 1.5 (0.56,4.3) (0.51,3.9) | 2.7 (0.88,8.3) (0.68,7.2) | 2.3 (0.70,7.5) (0.52,6.3) |
| Past abortion | 2 | 1 (0.25,4) | 1 (0.06,16) | 0.72 (0.18,2.9) (0.12,2.3) | 0.84 (0.34,2.1) (0.31,1.9) | 0.82 (0.34,2.0) (0.31,1.9) | 0.76 (0.23,2.5) (0.18,2.1) | 0.75 (0.23,2.5) (0.18,2.1) |
| Hydramnios | 1 | 4 (1,16) | 4 (0.25,64) | 60 (5.7,635) (2.8,478) | 5.8 (1.6,21) (1.6,22) | 1.5 (0.41,5.7) (0.41,6.3) | 17 (1.5,190) (1.4,158) | 8.1 (0.36,179) (0.44,117) |
| Labour progress | 2 | 2 (0.5,8) | 2 (0.13,32) | 0.50 (0.06,3.9) (0.04,2.8) | 1.3 (0.45,3.5) (0.42,3.3) | 0.83 (0.29,2.4) (0.27,2.3) | 0.80 (0.17,3.8) (0.13,3.2) | 0.67 (0.14,3.2) (0.10,2.8) |
| PCA | 1 | 2 (0.5,8) | 2 (0.13,32) | 3.1 (0.33,29) (0.15,20) | 2.2 (0.71,7.1) (0.67,7.0) | 1.3 (0.38,4.2) (0.36,4.2) | 2.6 (0.43,15) (0.33,13) | 1.8 (0.27,12) (0.22,11) |
| No monitor | 3 | 2 (0.5,8) | 2 (0.13,32) | 1.2 (0.32,4.9) (0.35,5.9) | 1.8 (0.68,4.5) (0.70,4.7) | 1.2 (0.47,2.8) (0.48,3.0) | 1.5 (0.43,5.0) (0.46,5.6) | 1.2 (0.37,4.0) (0.40,4.6) |
| Twin, triplet | 3 | 4 (1,16) | 4 (0.25,64) | 8.2 (1.8,37) (1.5,33) | 5.1 (1.9,14) (1.8,14) | 2.4 (0.74,7.6) (0.74,7.7) | 6.6 (1.7,25) (1.5,23) | 5.2 (1.2,23) (1.0,20) |
| Public ward | 6 | 2 (0.5,8) | 2 (0.13,32) | 0.86 (0.26,2.9) (0.25,2.8) | 1.3 (0.56,3.0) (0.54,3.0) | 1.1 (0.46,2.5) (0.45,2.5) | 1.0 (0.36,3.0)B (0.34,2.9) | 1.0 (0.35,2.9) (0.34,2.8) |
| PROM | 1 | 2 (0.5,8) | 2 (0.13,32) | 0.54 (0.06,4.8) (0.03,3.2) | 1.3 (0.45,3.5) (0.41,3.3) | 0.88 (0.30,2.6) (0.28,2.5) | 0.81 (0.17,3.9) (0.12,3.3) | 0.68 (0.13,3.5) (0.10,2.9) |
| Malpresentation | 3 | 4 (1,16) | 4 (0.25,64) | 3.9 (0.88,17) (0.73,15) | 3.9 (1.5,10.0) (1.4,9.8) | 1.9 (0.65,5.6) (0.64,5.5) | 3.8 (1.1,13) (0.93,13) | 3.0 (0.75,12) (0.66,11) |

ML: maximum likelihood; MCMC: Markov-chain Monte Carlo; PCA: placental or cord anomaly. PROM: premature rupture of membranes. Variables are indicators except early age (0 = 20+, 1 = 15-19, 2 = under 15), gestational age (0 = no, 1 = 36–38 weeks, 2 = 33–36 weeks; under 33 weeks excluded), isoimmunization (0 = no, 1 = Rh, 2 = ABO), labour progress (0 = no, 0.33 = prolonged, 0.67 = protracted, 1 = arrested) and past abortion (0 = none, 1 = 1, 2 = 2+). The intercept has a prior with 95% limits of ln(0.0001), ln(0.005). Syntax to reproduce this table is available in the online appendix (Supplementary data are available at *IJE* online).

the sample intercept is distorted by design constraints such as case-control sampling. Similarly, in matched case-control studies in which the matching covariates are controlled by entering them in unconditional logistic regression, their sample relation to disease (as well as the intercept) will be distorted by the matched sampling; thus external prior information will not be applicable and should not be used to set their priors without first accounting for the matching.

All methods require attention to convergence and collinearity problems, as well as the realism of the model (and priors, if Bayesian). Maximum-likelihood, the default in all commercial software, requires that the model information in the data is sufficient for large-sample approximations to work; failure is indicated by 'explosion' of estimates, as seen above for the hydramnios ML estimate and elsewhere.[17] Numerical studies indicate that the common criterion of at least 4 or 5 cases and 4 or 5 noncases per model coefficient suffices to avoid serious bias.[15,25] For data augmentation this criterion should be applied after including the number of subjects in the prior data but before rescaling. Thus, using prior data with at least A = 4 cases and M − A = 4 noncases per coefficient for all coefficients before rescaling by $S$, the criterion is automatically satisfied; this is why the Wald and profile-likelihood limits agree so well in the analyses using the priors in Table 1. It should be noted however that the criterion must be applied to A and M − A separately; simply requiring M to be large is inadequate, as can be seen in examples.[17] The criterion also does not take account of collinearities, although use of independent priors with A and M − A over 4 (as above) will address the need for independent information on regressors.

MCMC does not require large samples but does require convergence of the Markov Chains to the posterior distribution, a condition that is difficult to verify with absolute assurance and may be less assured with small samples or weak identification.[26] Because data augmentation and MCMC rely on very different conditions for convergence, running both procedures and comparing their output may be a cautious and advisable approach to Bayesian analysis involving complex models. Although they have provided indistinguishable results in the foetal-monitoring and other examples involving conventional logistic and log-linear models, especially when using profile instead of Wald limits,[6–8] important discrepancies could occur with intrinsically nonlinear models or when the posterior is multimodal or otherwise complex in shape. Thus, any conflict needs close investigation to determine which method (if either) is performing adequately.

## Supplementary Data

Supplementary Data are available at *IJE* online.

## Acknowledgements

The Melbourne WHO Collaborating Centre for Reference and Research on Influenza is supported by the Australian Government Department of Health and Ageing.

**Conflict of interest:** None declared.

## References

[1] Box GEP. Sampling and Bayes inference in scientific modeling and robustness. *J R Stat Soc A* 1980;**143:** 383–430.
[2] Good IJ. The Bayesian influence. In Harper W, Hooker CA (eds). *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Vol 2. Boston: D. Reidel Publishing Company, 1976, pp. 125–174. Reprinted in Good IJ. *Good Thinking*. Minneapolis, MN: University of Minnesota Press, 1983, pp. 22–55.
[3] Greenland S. Multiple-bias modelling for analysis of observational data. *J R Stat Soc A* 2005;**168:**267–91.
[4] Greenland S. Bayesian perspectives for epidemiological research. I. Foundations and basic methods. *Int J Epidemiol* 2006;**35:**765–75.
[5] Leamer E. *Specification Searches*. New York: Wiley, 1978.
[6] Cole SR, Chu H, Greenland S, Hamra G, Richardson DB. Bayesian posterior distributions without Markov chains. *Am J Epidemiol* 2012;**175:**368–75.
[7] Greenland S. Putting background information about relative risks into conjugate prior distributions. *Biometrics* 2001;**57:**663–70.
[8] Greenland S. Generalized conjugate priors for Bayesian analysis of risk and survival regressions. *Biometrics* 2003; **59**:92–9.
[9] Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *Int J Epidemiol* 2007;**36:** 195–202.
[10] Greenland S. Prior data for non-normal priors. *Stat Med* 2007;**26:**3578–90.
[11] Greenland S. Relaxation penalties and priors for plausible modeling of nonidentified bias sources. *Stat Sci* 2009;**24:** 195–210.
[12] Werler MM, Ahrens KA, Bosco JL *et al*. Use of antiepileptic medications in pregnancy in relation to risks of birth defects. *Ann Epidemiol* 2011;**21:**842–50.
[13] Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press, 1975.
[14] Good IJ. *The Estimation of Probabilities; an Essay on Modern Bayesian Methods*. Cambridge, MA: MIT Press, 1965.
[15] Greenland S. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics* 2000;**1:**113–22.
[16] Greenland S, Schwartzbaum JA, Finkle WD. Problems due to small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol* 2000;**151:** 531–39.
[17] AbdelSalam H, Restrepo C, Tarity TD, Sangster W, Parvizi J. Predictors of intensive care unit admission after total joint arthroplasty. *J Arthroplasty* 2012;**27:** 720–25.

[18] Neutra RR, Fienberg SE, Greenland S, Friedman EA. Effect of fetal monitoring on neonatal death rates. *N Engl J Med* 1978;**299:**324–26.
[19] Greenland S. Bayesian perspectives for epidemiologic research. III. Bias analysis via missing-data methods. *Int J Epidemiol* 2009;**38:**1662–73.
[20] SAS Institute Inc. *SAS/STAT® 9.2 User's Guide*. Second edition Cary, NC: SAS Institute, 2009.
[21] Greenland S. 18. Introduction to Bayesian statistics. In Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott, Williams & Wilkins, 2008.
[22] Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *Am J Epidemiol* 2008;**167:**523–29. Corrigendum 2008; 167:1142.
[23] Greenland S. Simpson's paradox from adding constants in contingency tables as an example of Bayesian noncollapsibility. *Am Stat* 2010;**64:**340–44.
[24] Heinze G, Ploner M. *A SAS Macro, S-PLUS Library and R Package to Perform Logistic Regression Without Convergence Problems*. Vienna: Medical University of Vienna, Department of Medical Computer Sciences, 2004, Report No. 2/2004.
[25] Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;**49:**1373–79.
[26] Cowles MK, Carlin BP. Markov Chain Monte Carlo convergence diagnostics: A comparative review. *J Am Stat Assoc* 1996;**91:**883–904.

# Appendix

Appendix SAS software code for Bayesian unconditional logistic regression by data augmentation and MCMC. Full syntax with data set available in the Supplementary Appendices at *IJE* online. Data and syntax for conditional logistic, Poisson and Cox regressions are also available online.

```
/*Retrieve data file*/
data fm0;
     infile fm0;
     input death nonwhite teenages nullip gestage isoimm abort hydram dyslab
placord nomonit twint ward prerupt malpres;
     Const=1; M=1; H=0;
     run;
/*create prior data using the offset method*/
data prior;
      input death nonwhite teenages nullip gestage isoimm abort hydram dyslab placord nomonit twint ward
prerupt malpres Const M H;
datalines;
4 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 8 −0.693147181
4 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 8 −0.693147181
4 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 8 −0.693147181
4 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 8 −1.386294361
4 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 8 −0.693147181
4 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 8 0
4 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 8 −1.386294361
4 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 8 −0.693147181
4 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 8 −0.693147181
4 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 8 −0.693147181
4 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 8 −1.386294361
4 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 8 −0.693147181
4 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 8 −0.693147181
4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 8 −1.386294361
;
/*add 1/2 to each prior count*/
data prior1;
     set prior;
     death=death+0.5;
     M=2*death;
     run;
/*add the prior data to the actual data*/
```

```
data augdata1;
     set fm0 prior1;
     run;
/*compute posterior estimates*/
proc logistic data=augdata1;
     title Bayesian logistic regression using data augmentation;
     model death/M = const nonwhite–malpres / noint offset=H clodds=both;
     run;
/*create rescaled prior data, A=400*/
data prior2;
     set prior;
     s=10;
     death=death*s*s;
     M=M*s*s;
     array covars nonwhite teenages nullip gestage isoimm abort hydram dyslab placord nomonit twint ward
prerupt malpres Const H;
     do i=1 to dim(covars);
       covars(i)= covars(i)/s;
       end;
     drop s i;
     run;
/*add the prior data to the actual data*/
data augdata2;
     set fm0 prior2;
     run;
/*compute posterior estimates*/
proc logistic data=augdata2;
     title Bayesian logistic regression using data augmentation, rescaled prior;
     model death/M = const nonwhite–malpres / noint offset=H clodds=both;
     run;
/*comparison with MCMC (Bayes statement in GENMOD*/
/*create dataset containing the prior variances and means*/
data prior3;
     input _type_ $ Intercept nonwhite teenages nullip gestage isoimm abort hydram dyslab placord nomonit
twint ward prerupt malpres;
datalines;
Var 1.0 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5
Mean −7.254479892 0.693147181 0.693147181 0.693147181 1.386294361 0.693147181 0
1.386294361 0.693147181 0.693147181 0.693147181 1.386294361 0.693147181 0.693147181 1.386294361
;
run;
/*compute MCMC posterior estimates*/
proc genmod data=fm0 descending;
     title Bayesian (MCMC) logistic regression;
     model death = nonwhite–malpres / dist=binomial link=logit lrci;
     bayes coeffprior=normal(input=prior3) stats(percent=2.5 50 97.5)
/* seed included to allow replication; not advised for actual analyses: */
     seed=1234 diagnostics=all plots=all;
     ods output PostSummaries=sum PostIntervals=int;
     run;
/*exponentiate the coefficients*/
data postmcmc;
     set sum;
     postll=exp(P2_5);
     postmed=exp(P50);
     postul=exp(P97_5);
     run;
proc print data= postmcmc;
     title MCMC posterior median and 95% limits;
     var parameter postmed postll postul;
     run;
```