INTRODUCTION

There is a growing literature on accounting for uncertainties in epidemiologic studies by using prior distributions (priors) for parameters that govern bias.1–26 Bayesian methods are a natural approach to use of priors, but epidemiologists have gravitated to more informal simulation methods. I here provide one approach to Bayesian uncertainty assessment, focussing on data priors in order to recast bias analysis as a missing-data problem. This approach shows how bias analysis is complementary to validation-study and two-phase (two-stage) analyses, and how all data and sources of error can be analysed simultaneously with missing-data methods.

I will assume the reader is familiar with bias simulation and basic Bayesian ideas as described in this journal8,14,27–29 and subsequent book chapters.22,30 The methods used here are partial-Bayes or semi-Bayes in which explicit priors are used for some, but not all, parameters.10,30–32 The Gauss code, libraries and output used for the main analysis below are available at http://www.ph.ucla.edu/epi/faculty/greenland/index.htm.

A case–control study with exposure misclassification

Conventional analyses

Table 1 presents a study of the relation of sudden infant death syndrome (SIDS) to maternal antibiotic use during pregnancy.33 Given the rarity of SIDS, the population risk ratio comparing the exposed with the unexposed (X=1 vs X=0) is approximated by the corresponding population odds ratio ORXY. If we ignore biases, we may take this odds ratio or its log, \( \beta = \ln(OR_{XY}) \), as the target parameter. The usual maximum-likelihood estimate (MLE) of \( OR_{XY} = e^\beta \) is the sample odds ratio, \( e^\beta = 173(663)/134(602) = 1.42 \), with standard error for \( \hat{\beta} \) of \( (1/173 + 1/602 + 1/134 + 1/663)^{1/2} = 0.128 \) and 95% confidence limits
Consider next a prior for the odds ratio. At the time of the study only weak speculations could be made about the size or direction of the association; antibiotics might be associated with elevated risk (marking effects of an infection on the fetus, or via direct effects), or with reduced risk (by reducing presence of infectious agents). Nonetheless, ‘large’ odds ratios seemed unlikely, because such effects would have led to SIDS ‘outbreaks’ in conditions with high antibiotic usage. Also, prenatal antibiotic use in the USA had climbed to ~20% over the preceding four decades, yet the SIDS rate remained a fraction of a percent. Thus, one reasonable starting point would place 2:1 odds on OR_{XY} between 1/2 and 2, and 95% probability on OR_{XY} between 1/4 and 4. Among the priors that would yield these bets would be a normal(0, 1/2) prior on \( \logit(\theta) \) and \( \sigma^2 = 1/2 \). But the prior mean is \( \exp(0.341, 1.96 - 0.128) = 1.10, 1.80 \), which has 95% probability on OR_{XY} \( \approx 0.352 \), and approximate posterior median of \( \exp(0.341) = 1.41 \). The prior sample \( \exp(0.341 \pm 1.96 \cdot 0.128) = 1.10, 1.80 \), which are barely distinct from the conventional results. The similarity arises because the data information about \( \theta \) is \( 1/0.128^2 = 61 \) but the prior information is only \( 1/0.500 = 2 \) (giving the sample estimate 30 times the weight of the prior mean), and because the prior and sample means are not far apart. Thus, the prior appears weak compared with the data, and the frequentist and partial-Bayes results are almost identical.

Essentially, the same results are obtained by using a fully Bayesian analysis with a weak prior on \( p = Pr(X = 1 | Y = 0) \approx Pr(X = 1) \), the prevalence of antibiotic use in the source population. For example, the OR_{XY} \( \approx 0.352 \) and variance of 1, which has 95% prior limits for \( p \) of 3.4 and 64.0%.

A misclassification problem
In the above, \( X \) represents only mother’s report of antibiotic use. Let \( T \) be the indicator of actual (true) antibiotic use. There is no doubt that mistaken reports \( (X \neq T) \) occur. Moreover, recall bias seems likely, with false positives \( (X > T) \) more frequent among cases and false negatives \( (X < T) \) more frequent among controls. The resulting causal structure is shown in Figure 1a; the parentheses around \( T \) indicate it is unobserved.

Let \( A_{xy} \) be the actual (but unobserved) count at \( T = t, X = x, Y = y \), let \( E_{ty} \equiv E(A_{ty}) \) be the expected count, and let a ‘+’ subscript indicate summation over the subscript. The problem can then be restated as: we observe only the \( XY \) margin (the counts in Table 1) and get an estimate \( A_{+1} = 673/A_{+0} = 663 \) of the marginal OR_{XY} \( E_{+1} = E_{+0} = E_{+1} \). But the odds ratio of substantive interest (i.e. the real target parameter) is the marginal OR_{XY} \( = E_{+1} / E_{+0} \). With no measurement of \( T \); data on \( T \) are missing for everyone; hence OR_{XY} cannot be computed from the observed \( XY \) counts.

Non-identified classification parameters
To what degree can the interval for OR_{XY} be taken as an inference about OR_{XY}? The answer requires construction of estimates of OR_{XY} for comparison. To estimate OR_{XY} we need additional information, such as prior distributions, subjects with data on \( T \), or both, that allows us to connect the marginal \( XY \) observations to OR_{XY}. Examples include information on predicting \( T \) from \( XY \), i.e. information on the predictive values \( \pi_{xy} \equiv Pr(T = t | X = x, Y = y) = E_{ty} / E_{xy} \). Because \( \pi_{xy} = 1 - \pi_{1xy} \) there are only four distinct classification parameters, which may be taken as the \( \pi_{1xy} \). Knowing the \( \pi_{1xy} \) would allow us to impute \( T \) in the data, as shown in Table 2. Unfortunately, by themselves, the \( XY \) data in Table 1 tell us nothing.

Table 1 Data from case–control study of SIDS

<table>
<thead>
<tr>
<th></th>
<th>( \text{X=1} )</th>
<th>( \text{X=0} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y=1 )</td>
<td>( A_{+1} = 173 )</td>
<td>( A_{+0} = 602 )</td>
</tr>
<tr>
<td>( Y=0 )</td>
<td>( A_{+1} = 134 )</td>
<td>( A_{+0} = 663 )</td>
</tr>
</tbody>
</table>

(1) generates mean of logit(0.2) and variance of 1, which has 95% prior limits for \( p \) of 3.4 and 64.0%.
Table 2 Imputed complete-data table from SIDS study

<table>
<thead>
<tr>
<th></th>
<th>$X = 1$</th>
<th>$X = 0$</th>
<th>$Y = 1$</th>
<th>$Y = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = 1$</td>
<td>173</td>
<td>134</td>
<td>602</td>
<td>663</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>173</td>
<td>134</td>
<td>602</td>
<td>663</td>
</tr>
<tr>
<td>Totals</td>
<td>173</td>
<td>134</td>
<td>602</td>
<td>663</td>
</tr>
</tbody>
</table>

$T$ indicates actual antibiotic use during pregnancy.

about these parameters, i.e. the $\pi_{xy}$ are not identified by those data.

One must impose assumptions about the classification process to say anything about the target OR$_{xy}$ based on the $XY$ data. Yet, absent such assumptions, default reporting offers the OR$_{xy}$ estimate (1.42, 95% CLs 1.11–1.83) as an estimate of OR$_{xy}$. This is exactly the answer one gets from assuming no error ($X = T$). It is also the answer from a partial-Bayes analysis using a single-point prior for the $\pi_{xy}$ that assigns 100% probability to $\pi_{11} = \pi_{00} = 1$. This is a contextually absurd prior that no one holds. In other words, when authors present conventional estimates as ‘effects’, they assume with certainty that a ridiculous extreme holds. Such presentations take no account of the actual uncertainty or prior information about the $\pi_{xy}$, which is vague but at least bounds the $\pi_{xy}$ away from 0 and 1. This criticism also applies to the conventional Bayesian results, which are based on the same absurd point priors for the $\pi_{xy}$.

Bias analysis as regression with missing covariates

To recast the problem in a conventional regression framework, we first represent the unobserved $TXY$ counts in terms of a log-linear model for the case ($Y = 1$) and control ($Y = 0$) counts,

$$E_{xy} = \exp\left(\beta_0 + \beta_T + \beta_X + \beta_Y + \beta_{TX}X + \beta_{TY}Y + \beta_{TXY}XY\right)$$

(1)

This model uses as outcomes the data counts, rather than proportions. Nonetheless, it implies logistic models for proportions, including the predictive-value model

$$\pi_{xy} = \frac{E_{xy}}{E_{xy} + E_{0xy}} = \expit\left(\beta_T + \beta_{TXY}X + \beta_{TY}Y + \beta_{TXY}XY\right)$$

(2)

where expit($u$) $\equiv e^u/(1 + e^u)$ is the logistic transform. This derived model shows how to impute the missing $T$ from the observed $X$ and $Y$ and the beta parameters.

**A plausible prior specification**

Working with the $T$-predictive coefficients $\beta_T$, $\beta_{TX}$, $\beta_{TY}$, $\beta_{TXY}$ instead of predictive values $\pi_{xy}$ cases specification of priors. Table 3 gives one set of normal priors for the logistic coefficients ($\beta_T$, $\beta_{TX}$, $\beta_{TY}$, $\beta_{TXY}$). This type of independent-normal prior structure for predictive-value coefficients has been used for other poorly understood associations, and is plausible the present example. Although the priors are only historical and non-specific, they show the impact of replacing the absurd single-point priors implicit in the conventional analyses by weaker priors that meet the following plausibility considerations.

To begin, expit($\beta_T$) = $\pi_{100}$ = Pr($T = 1 | X = 0, Y = 0$) is the probability that a ‘test negative’ ($X = 0$) in a non-case is erroneous. For an exposure with an expected prevalence well below 50% (such as antibiotic use in unselected pregnancies) and a reasonably specific test, we would concentrate our prior distribution for $\pi_{100}$ well below 0.5, which forces the prior for $\beta_T$ to be well below 0. Let $\phi_{0y} = Pr(X = x | T = t, Y = y)$. The $\phi_{1y}$ are the sensitivities and the $\phi_{00}$ are the specificities for $X$ as a measure of $T$, while $\phi_{10}$ and $\phi_{01}$ are the false-negative and false-positive probabilities. One summary of accuracy is the $TX$ or receiver operating-characteristic (ROC) odds ratio, which is the true-positive odds over false-positive odds:

$$\text{OR}_{TX} = \frac{\phi_{10}}{\phi_{01}} = \frac{(\pi_{10})}{(\pi_{01})} = \exp(\beta_T + \beta_{TXY})$$

For a worthless measurement $X$ is independent of $T$ given $Y$, so $\text{OR}_{TX} = 1$ and $\beta_T = \beta_{TXY} = 0$. For sensitivity 0.6 and specificity 0.8 among non-cases, $\text{OR}_{TX0} = \exp(\beta_{TX}) = (0.6/0.4)/(0.2/0.8) = 6$; 0.6 and 0.9 yield (0.6/0.4)/(0.1/0.9) = 13.5; 0.8 and 0.8 yield (0.8/0.2)/(0.2/0.8) = 16; and 0.8 and 0.9 yield (0.8/0.2)/(0.1/0.9) = 36. Thus, even for a mediocre measurement the TX odds ratio among non-cases, $\text{OR}_{TX0} = \exp(\beta_{TX})$, should be high and a prior for $\beta_{TXY}$ should be distributed well above zero.

The change in $\text{OR}_{TX}$ moving from non-cases to cases is $\text{OR}_{TX1}/\text{OR}_{TX0} = \exp(\beta_{TXY})$, hence $\beta_{TXY} > 0$ if cases have more accurate recall than non-cases, $\beta_{TXY} < 0$ if vice versa. It is difficult to judge which is more likely. Under a recall-bias hypothesis, cases had better sensitivity (higher true-positive odds) but poorer specificity (higher false-positive odds), which have opposing effects on $\text{OR}_{TX}$. Thus, a prior for $\beta_{TXY}$ centred at zero appears reasonable.

A common assumption is that the misclassification is non-differential, i.e., that $X$ and $Y$ are independent.
given $T$, or equivalently, equal sensitivity and specificity across $Y$. Because $\beta_{XY}$ and $\beta_{XY} + \beta_{XX}$ are the $XY$ log odds ratios in the $T=0$ and $T=1$ strata, under non-differentiality we have $\beta_{XY} = \beta_{XX} = 0$, giving $OR_{TY} = \exp(\beta_{TY})$. When $X$ is determined with knowledge of $Y$, as here, non-differentiality is not a justifiable assumption. Nonetheless, if we expect limited impact of $Y$ on $X$, our priors on $\beta_{XY}$ and $\beta_{XX}$ would not spread far from zero. In that case we obtain only small departures of $\exp(\beta_{TY})$ from $OR_{TY}$, which suggests a prior for $\exp(\beta_{TY})$ similar to that for $OR_{TY}$, perhaps expanded slightly to allow for uncertainty about $\beta_{XY}$ and $\beta_{TY}$.

The other parameters may be left with no explicit prior. For $\beta_0$ and $\beta_1$ the absence of a prior is dictated by the fact that these are both primarily determined by the study design, with $\beta_0$ reflecting the size of the study and $\beta_1$ reflecting the case–control ratio, subject to uncertain selection and refusal forces. For $\beta_X$ and $\beta_{XY}$ the absence of a prior is primarily a matter of convenience, recognizing that the large $XY$ margins (the $A_{XY}$) and weak background information will render their priors of secondary importance (although results using these priors will be given later).

The resulting partial-Bayesian analysis is a direct extension of the conventional analysis to allow for imperfect knowledge about parameters (the $\pi_{TY}$) that were previously assumed known but are completely undetermined (not identified) from the data alone. It has a frequentist interpretation as a penalized-likelihood analysis in which penalty functions (equal to $-2$ times the log priors) replace the unsatisfactory point constraints $\pi_{11Y} = \pi_{00Y} = 1$.10 The general steps are as follows.

**Translating the priors into data**

The data in Table 1, the $T$-predictive model (2), and the priors in Table 3 compose the input to a partial-Bayesian analysis. One may compute posterior intervals from these inputs in many ways, such as posterior sampling (e.g. Markov-Chain Monte-Carlo)17–40 or Monte-Carlo sensitivity analysis (MCSA).9,22,24

Another approach instead uses conventional missing-data methods on a data set that treats the actual data records as records missing $T$, and then adds a complete record for each parameter ($\beta_T$, $\beta_{XY}$, $\beta_{XX}$) in the regression of $T$ on $X$ and $Y$. The added record for a given parameter is derived from its normal ($\mu$, $\tau^2$) prior to make the likelihood contribution from the record proportional to the prior for the parameter.6,10,41 The general steps are as follows.

(i) Reformat the actual data as a data set with $T$ a missing covariate: add a column for $T$ to the data set and enter the missing-value code in that column in the actual records.

(ii) Augment the actual data records with prior data records representing the desired prior distributions.

(iii) Conduct an analysis of the augmented (actual plus prior) data using a valid, reasonably efficient method for handling missing-at-random (MAR) data.

The Appendix and citations describe more general data priors for logistic, log-linear, and survival regression.10,41–44

Consider the log-linear model (1) for observed counts and an independent normal ($\mu$, $\tau^2$) prior on a model coefficient $\beta_Z$ where $Z = T$, $TX$, $TY$ or $XX$. We may force this prior on $\beta_Z$ by appending the following prior-data record to the actual data set.

(i) The number of events $A$ in the record is $s^2/\tau^2$, where $s$ is a scaling factor. Normality increases as $s$ increases; $s = 30$ assures the approximation is reasonable for Poisson data.28,43

(ii) $Z = 1/s$ in the record. This is the value of $Z$ that makes the variance $\tau^2$ of $\beta_Z$ estimated from the record equal to $\tau^2 = 1/AZ^2$ when $A = s^2/\tau^2$.41,43

(iii) The person-time for the record is $N = A/\exp(\mu/s) = s^2\exp(-\mu/s)/\tau^2$. This is the value for $N$ that makes the mean $\mu$ of $\beta_Z$ estimated from the record equal to $\mu = s\ln(A/N)$. Equivalently, it is the solution for $\mu$ of the rate equation $A/N = \exp(\mu Z)$ when $Z = 1/s$ and $A = s^2/\tau^2$. Note that if $\mu = 0$ then $N = A$.

(iv) All other covariates in the prior record are set to zero, including the constant

For the actual-data records we set the person-time to $N = 1$ unless the counts arise from actual person-time, in which case that time is used. An alternative method that uses binomial logistic regression to model the counts is described in the appendix.

The steps just outlined correspond to the ‘offset method’ in Greenland,28 except that treating $\exp(H)$ as the person-time is equivalent to entering the offset $H$ in the regression. Because the method requires setting the constant to zero in the prior records, one must suppress the automatic intercept option in the regression program (via an option like ‘no constant’ or ‘intercept = 0’). In its place we add to the entire data set an actual-data indicator $D$ defined as $D = 1$ for the actual-data records, 0 for the added (prior) records. The coefficient of $D$ is the intercept $\beta_0$.28 If we have a normal($\mu$, $\tau^2$) prior for $\beta_0$ it may be represented as an additional record with $A = s^2/\tau^2$ events, $N = A/\exp(\mu/s)$, $D = 1/s$ and all other covariates zero.

Applying steps 1–4 with the priors in Table 3, we obtain four prior records, each with $A = 900/\tau^2$ and zeros for all covariates except one.

(i) For $\beta_T \sim$ normal($\mu$, 0.16) with $\mu = \logit(0.1) = -\ln(9)$ we get $A = 900/0.16 = 5625$, $N = 5625/\exp[\ln(9)/30] \approx 5228$, $T = 1/30$.

(ii) For $\beta_{XY} \sim$ normal($\mu$, 0.25) with $\mu = \ln(13.5)$ we get $A = 900/0.25 = 3600$, $N = 3600/\exp[\ln(13.5)/30] \approx 3301$, $TX = 1/30$.

(iii) For $\beta_{TY} \sim$ normal(0, 0.5) we get $A = 900/0.5 = 1800$, $N = 1800/\exp(0/30) = 1800$, $TY = 1/30$. 
(iv) For $\beta_{TXY} \sim \text{normal}(0, 0.125)$ we get $A = 900/0.125 = 7200$, $N = 7200/\exp(0/30) = 7200$, $TXY = 1/30$.

Note that the prior records for product terms contain logically impossible covariate combinations, e.g., $TX$ in an actual record must be 0 if $T$ or $X$ is 0, but the prior record for $\beta_{TX}$ has 0 for all covariates except $TX$. This oddity shows that the coding scheme is a way of making the program create a likelihood contribution proportional to a normal prior for $\beta_{TX}$ making the program create a likelihood contribution oddity shows that the coding scheme is a way of making the program create a likelihood contribution.

Table 4 shows the four actual-data records, composed of four Poisson $XY$ counts with $T$ missing; these have $N = 1$ and $D = 1$. It also shows the four complete-data prior records derived from converting the priors for the four non-identified parameters ($\beta_{T}, \beta_{TX}, \beta_{TY}, \beta_{TXY}$) into four complete records. Priors for other coefficients such as $\beta_x$ and $\beta_{XY}$ could be added as further records.

Having replaced the constant with an indicator $D = 1$ for actual data, 0 for prior data, let $F_{dxy}$ be the expected count when $D = d$, $T = t$, $X = x$, $Y = y$. The model for the augmented-count data is now

$$F_{dxy} = \exp \left( \beta_0d + \beta_Tt + \beta_XX + \beta_TY + \beta_{TX}tx + \beta_{TY}ty + \beta_{TXY}txy \right)$$  

and the expected actual-data counts are $F_{1xy} = E_{1xy}$. This model differs from the original (1) only in that $\beta_0$ multiplies the actual-data indicator instead of standing alone.

Computation

We may fit model (3) to the augmented data using any efficient method for Poisson regression on partially missing covariates. With approximately normal priors, MLEs from conventional missing-data software will be approximate posterior medians, and the accompanying confidence limits will be approximate posterior limits at the same percentage. This approach (coupled with simulation intervals) is illustrated by the program at the webpage mentioned earlier and used to generate the results below. Similar results can be obtained using multiple imputation (MI) with at least 10 imputations. In this context, MI can be viewed as a natural extension of the record-level imputation method, and corresponds to MI for measurement-error correction with prior data in place of validation data.

Putting the coefficient estimates in model 1, we obtain fitted actual-data counts $E_{1xy}^T$ that can be summed over $X$ to get a $TY$ table with counts $E_{1xy}^T$. The approximate posterior median for the target $OR_{TY}$ is then $OR_{TY} = E_{1_{+1}}^T/\exp(0/30)/E_{1_{+0}}^{T+0}$; this is the point at which we would give 1:1 odds for $OR_{TY}$ being above or below, under the data, model and priors. Using MI estimates, $OR_{TY} = 1.19$ in the example.

Posterior limits for $OR_{TY}$ can be approximated by confidence limits derived from the augmented data, which may be easily computed by resampling (bootstrap) methods. Most rapid is to resample the coefficients from their approximating multivariate normal distribution using the point estimates and estimated covariance matrix, then re-calculate the expected cells to get a new $OR_{TY}$; here, 250,000 samplings give 95% confidence (and posterior) limits of 0.41, 3.5. Slightly more accurate limits of 0.37, 3.4 derive from resampling of all (actual plus prior) data counts 250,000 times (using the counts in column ‘A’ of Table 4 as Poisson means), refitting the model and recalculating OR_{TY}; the same limits are obtained by exact posterior simulation and from MCSA.

All methods show that the precision of the conventional frequentist and Bayesian results given earlier (95% limits for $OR_{TY}$ of 1.1, 1.8) is due to the absurdly precise assumptions (single-point priors) that both use for the predictive values $\pi_{xy}$ i.e. that $\pi_{11} = \pi_{00} = 1$. Results from contextually reasonable

Table 4 Data records used to replicate Bayesian analysis via penalized likelihood

<table>
<thead>
<tr>
<th>Actual-data records</th>
<th>A</th>
<th>N</th>
<th>D</th>
<th>T</th>
<th>X</th>
<th>Y</th>
<th>TX</th>
<th>TY</th>
<th>XY</th>
<th>TXY</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X = 1$, $Y = 1$</td>
<td>173</td>
<td>1</td>
<td>1</td>
<td>Mis.</td>
<td>1</td>
<td>1</td>
<td>Mis.</td>
<td>1</td>
<td>Mis.</td>
<td>1</td>
</tr>
<tr>
<td>$X = 0$, $Y = 1$</td>
<td>602</td>
<td>1</td>
<td>1</td>
<td>Mis.</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Mis.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X = 1$, $Y = 0$</td>
<td>134</td>
<td>1</td>
<td>1</td>
<td>Mis.</td>
<td>1</td>
<td>0</td>
<td>Mis.</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X = 0$, $Y = 0$</td>
<td>663</td>
<td>1</td>
<td>1</td>
<td>Mis.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prior-data records</th>
<th>$\beta_T$</th>
<th>$\beta_{TX}$</th>
<th>$\beta_{TY}$</th>
<th>$\beta_{TXY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_T$</td>
<td>5625</td>
<td>3600</td>
<td>1800</td>
<td>7200</td>
</tr>
<tr>
<td>$\beta_{TX}$</td>
<td>5228</td>
<td>3301</td>
<td>1800</td>
<td>7200</td>
</tr>
<tr>
<td>$\beta_{TY}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_{TXY}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

First four records are the actual data from Table 1, last four records are the priors from Table 3.

$A =$ event count.

$N =$ person-time.

Mis. $=$ missing.
priors show instead that the data add little information about OR\textsubscript{TY} beyond that conveyed by the priors, whether those priors are hidden (as in conventional analysis) or explicit (as in bias analysis).

**Relation to Monte-Carlo sensitivity analysis**

Because priors were placed only on non-identified parameters, the resampling limits are the same as those obtained from MCSA in which at each cycle

(i) all counts are resampled from Table 1,

(ii) \( \beta_T, \beta_{TX}, \beta_{TY} \) and \( \beta_{TXY} \) are drawn from their priors and used to compute the \( \pi_{txy} \),

(iii) these \( \pi_{txy} \) are multiplied against the resampled counts to get a simulated \( TXY \) table and

(iv) the table is collapsed over \( X \) and \( OR_{TY} \) is computed from the resulting \( TY \) table.\(^9,10\)

MCSA simulation does not, however, extend as easily to situations in which identified parameters are given priors, whereas Bayesian computation need only add additional records for those priors.

As an example, note that non-differentiality (independence of the classification error \( X-T \) and \( Y \)) corresponds to \( \beta_{XY}=\beta_{TXY}=0 \). This assumption is unreasonable (due to the potential for recall bias), but it is reasonable to constrain \( \beta_{XY} \) with a prior (\( \beta_{TXY} \) already has a prior). Suppose we had a normal(0, 1/4) prior on \( \beta_{TXY} \), which is induced by adding a record with \( A=900/0.25=3600, \ XY=1/30, \) all other covariates 0 and has 95% limits of 0.375, 2.66. With this prior in addition to the others, \( OR_{TY} \) becomes 1.4 with approximate 95% posterior limits of 0.62, 3.3, hardly more precise than without the prior (the upward shift is due to the constraint on differentiality imposed by the \( \beta_{TY} \) prior). But if instead we assume non-differentiality (\( \beta_{XY}=\beta_{TXY}=0 \)), \( OR_{TY} \) would be 1.8 with 95% posterior limits of 1.1, 2.8, displaying the excessive certainty about \( OR_{TY} \) produced by excessive certainty about non-differentiality.

**Priors for sensitivities and specificities**

Priors on sensitivities \( \phi_{t1y} \) and specificities \( \phi_{00y} \) are sufficient for identifying \( OR_{TY} \) but can be difficult to specify realistically because of dependencies among them.\(^{14,22}\) For example, case and non-case values \( \phi_{111} \) and \( \phi_{110} \) must be highly correlated even when differential (unequal), whereas sensitivities and specificities have sources of positive as well as negative correlation.\(^{22}\) Priors on the \( \phi_{txy} \) can also cause difficulty for MCSA when their range extends beyond values compatible with the data.\(^{14,22}\)

An alternative that minimizes these problems is to use priors on those coefficients in (1) that determine the \( \phi_{txy} \) via the logistic regression

\[
\phi_{t1y} = Pr(X=1|T=t, Y=y) = \frac{e_0}{e_{1+y}}, \quad \phi_{00y} = \expit(\beta_X + \beta_{TY} + \beta_{TXY} Y).
\]

Independent priors for the coefficients (\( \beta_X, \beta_{TX}, \beta_{TY}, \beta_{TXY} \)) will create prior dependencies among the \( \phi_{txy} \) as desired. Priors may be plausibly specified by noting that \( \expit(\beta_X) = Pr(X=1|T=0, Y=0) = P_{F0} \) is the false-positive probability among non-cases; the \( OR_{TY} \) are \( \phi_{11y}/\phi_{10y}/\phi_{01y}/\phi_{00y} = \exp(\beta_{TX} + \beta_{TXY} Y) \); and that \( OR_{TY} = \exp(\beta_{TY} + \beta_{TXY} Y) \), the odds ratio relating \( X \) to \( Y \) when \( T=t \), is 1 under non-differentiality.

For identification, it suffices to specify priors on \( \beta_X, \beta_{TX}, \beta_{TY} \) and \( \beta_{TXY} \). Using the priors specified earlier for \( \beta_X, \beta_{TY}, \beta_{TXY} \) along with the same prior for \( \beta_Y \) as used above for \( \beta_T \), \( OR_{TY} \) becomes 2.1 with approximate 95% posterior limits of 0.45 and 11, which is even less informative than the results using priors for \( \beta_T, \beta_{TX}, \beta_{TY} \) and \( \beta_{TXY} \). The difference is due almost entirely to replacing the original(0, 1/2) prior for \( \beta_{TY} \) by the normal(0, 1/4) prior for \( \beta_{TY} \), reflecting that the information about \( OR_{TY} \) conveyed by the \( \beta_{TY} \) prior is much more indirect than the information conveyed by the \( \beta_{TY} \) prior. Again, the upward shift is due to the constraint on differentiality imposed by the \( \beta_{TY} \) prior.

**Validation studies**

After allowing for uncertainties about misclassification, the SIDS data offer far less information about the target parameter than the conventional analysis makes it seem. One way to address this problem is to obtain \( T \) for some subjects, so that there are actual-data records with \( T \) present.

A validation sub-study

In the SIDS example, validation data were obtained by sampling the maternal medical records of the original study subjects.\(^{47}\) Table 5 separates the data from Table 1 into a \( T \)-known (validation or complete-data) stratum in the \( T=1 \) and \( T=0 \) rows, and a \( T \)-unknown (unvalidated or incomplete-data) stratum in the final row. Also shown are the estimates \( \hat{\pi}_{t1y}=C_{11y}/C_{++y} \) of \( \pi_{t1y} \), where \( C_{11y} \) is the number in the substudy with \( T=t, X=x, Y=y \). If validation selection is random then \( T \) is missing at random, and the MLE \( OR_{TY} \) of \( OR_{TY} \) simplifies to the odds ratio obtained by using the \( \hat{\pi}_{txy} \) from Table 5 in place of the \( \pi_{txy} \) in Table 2 to impute \( T \) where it is missing, followed by collapsing over \( X \):\(^{48}\)

\[
OR_{TY} = \frac{\hat{E}_{+1y}}{\hat{E}_{+0y}} \frac{\hat{E}_{1+y}}{\hat{E}_{+1y}}
\]

where

\[
\hat{E}_{1+y} = \sum_x \hat{\pi}_{txy} A_{+xy}
\]

In the example this MLE is 1.21 with Wald 95% CLs of 0.79 and 1.87.\(^{48,49}\) If we impose a constraint such as non-differentiality, the MLE need no longer have closed form, but may be replaced by any nearly efficient closed-form estimator.\(^{59}\)
As with the bias analysis, the validation-data analysis shows that the apparent precision of the conventional results is illusory. The change is far less dramatic, however, because the validation data are given far more weight than the prior. The prior and validation data can be further compared and combined by replacing the four actual-data records in Table 5 with the 12 actual-data records that result from the partial measurement of $T$. Table 6 shows these records; the four prior records in Table 4 would be appended to the data file, along with columns for $N$ and $D$. Fitting the earlier model for $F_{txy}$ to these 16 records, we obtain an approximate posterior median of 1.20 and 95% limits for OR$_{xy}$ of 0.81, 1.77, close to the results without the prior. Thus the prior, while informative, is much less informative than the validation data assuming MAR.

### Fundamental limitations of validation studies

The high weight given the validation data in the example is a consequence of assuming that $T$ is error free and MAR (i.e. validated randomly within levels of $X$ and $Y$). A validation sub-study provides estimates of the coefficients ($\hat{\beta}_T$, $\hat{\beta}_{TX}$, $\hat{\beta}_{TY}$, $\hat{\beta}_{TXY}$) in the logistic regression of $T$ on $X,Y$, but these estimates are valid only to the extent that we properly model the sampling and measurement of $T$. If we assume that $T$ is error free, unrelated to the validation-sampling probability and the model is correct, it follows that we can ignore these sampling and measurement mechanisms.

If, however, we delve deeper into the context, we often find that this simple validation-data model cannot be correct. Validation may require more cooperation from subjects or physicians and so is not likely to be completely random, and validation data will

### Table 5 SIDS data separated into validated strata (medical record examined) and unvalidated strata

<table>
<thead>
<tr>
<th></th>
<th>$Y = 1$</th>
<th>$Y = 0$</th>
<th>$X = 1$</th>
<th>$X = 0$</th>
<th>$X = 1$</th>
<th>$X = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 1$</td>
<td>29</td>
<td>17</td>
<td>21</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 0$</td>
<td>22</td>
<td>143</td>
<td>12</td>
<td>168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>51</td>
<td>160</td>
<td>33</td>
<td>184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\pi}_{xy}$</td>
<td>0.569</td>
<td>0.106</td>
<td>0.636</td>
<td>0.087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unvalidated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T$ missing</td>
<td>122</td>
<td>442</td>
<td>101</td>
<td>479</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imputed counts (for $T = 1$, $\hat{\pi}<em>{xy}$ times unvalidated count; for $T = 0$, $\hat{\pi}</em>{xy}$ times unvalidated count)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 1$</td>
<td>73.2</td>
<td>44.2</td>
<td>60.6</td>
<td>47.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T = 0$</td>
<td>48.8</td>
<td>397.8</td>
<td>40.4</td>
<td>431.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$W = 1$ if record shows use, $T = 0$ if not.

### Table 6 Actual-data records incorporating validation data from Table 5*

<table>
<thead>
<tr>
<th>$A$</th>
<th>$T$</th>
<th>$X$</th>
<th>$Y$</th>
<th>$TX$</th>
<th>$TY$</th>
<th>$XY$</th>
<th>$TXY$</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>143</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>168</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>122</td>
<td>Mis.</td>
<td>1</td>
<td>1</td>
<td>Mis.</td>
<td>Mis.</td>
<td>1</td>
<td>Mis.</td>
</tr>
<tr>
<td>442</td>
<td>Mis.</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Mis.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>101</td>
<td>Mis.</td>
<td>1</td>
<td>0</td>
<td>Mis.</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>479</td>
<td>Mis.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*These 12 records replace the four actual-data records in Table 4; all have $N = 1$, $D = 1$. 
often contain errors. In the example, the medical record only indicates antibiotic prescription, not use, so that at best \( T \) is an ‘intent-to-treat’ variable and the observed association is a prescribing effect.

An analysis that accounted for this problem would require relabelling \( T \) as an alternative measurement. The true value would be reintroduced as an unobserved ‘usage’ variable \( U \), which would have only missing values in the actual data in Table 6, and would be zero in all the \( T \)-prior data in Table 5. Identification would now require setting a prior on the distribution of \( U \) given \( T, X, Y \). If this distribution were vague, the posterior interval for \( \text{OR}_{TY} \) would remain very wide after obtaining the ‘validation’ \((T, X, Y)\) data, illustrating how the apparent precision of the validation analysis is purchased only by using extremely informative priors such as \( T = U \).

When precise priors or models are accepted as facts, agreeable to everyone (as when \( T \) indicates only ‘male’/’female’), the resulting precision of interval estimates should be agreeable to everyone. But in nutritional, occupational and environment studies even our best measurements are far from error free. Hence all precise-looking results must ultimately depend on strong non-identified assumptions (point priors) relating measurements to unidentified targets. Once we replace these assumptions with diffuse priors we see that all results depend on these priors in an unlimited fashion, no matter how many measurements are obtained. This sensitivity is illustrated in the example, which shows that any validation data can be rendered wholly non-informative by giving a non-informative prior to the conditional distribution of \( U \) given \( T, X \) and \( Y \).

Use of more measurements, as found in state-of-the-art methods,\(^5\) depends on point priors within more complex error structures. These structures can be much more plausible than conventional ones. Nonetheless, when the true time-varying exposure history can never be known without error, the use of point priors can never be justified empirically. Such use is instead a heuristic to deal with overwhelming complexity, and tends to produce overconfident formal statistics. This pitfall is an unavoidable limitation of formal statistical inference.

Uncontrolled confounders

Consider now a setting in which \( X \) rather than \( T \) is the exposure variable of interest and \( T \) is an unmeasured confounder of the effect of \( X \) on \( Y \), as in Figure 1b. Because we must consider the entire \( TXY \) distribution and \( T \) is unobserved, the regression models used for misclassification still apply. The target parameter shifts, however, becoming a summary of the pair of \( T \)-conditional \( XY \) odds ratios \( \text{OR}_{XY0} = \exp(\beta_{XY}) \) and \( \text{OR}_{XY1} = \exp(\beta_{XY} + \beta_{TXY}) \). It is usually assumed that these odds ratios are equal, which forces \( \beta_{TXY} = 0 \) and leaving \( \exp(\beta_{XY}) \) as the target. This assumption almost never has any firm empirical foundation and corresponds to a point prior (\( \beta_{TXY} = 0 \) with probability 1), although it may often have only minor impact on estimation of summaries.\(^5\)

Regardless, the Bayesian approach proceeds as in the misclassification example: the unmeasured confounder \( T \) is added as a new column in the data set with a missing value code for the actual records, and the priors relating \( T \) to the observed variables are added as new records. In two-phase studies \( T \) is measured on a subsample of subjects.\(^5\)\(^2\)\(^3\) Second-phase sampling corresponds to validation subsampling, and the resulting complete records may be entered into the analysis directly (analogously to Table 6).

Selection bias

Consider next a setting in which \( T \) is again the exposure variable of interest, but with \( X \) now a non-response (failure to select, locate or recruit) indicator in the study. Only subjects with \( X = 0 \) are observed. Figure 1c shows the structure (square brackets on \( X \) indicate observation is conditional on \( X \)). Again, the models used for misclassification are unchanged, and the target parameter is the same as in that case, \( \text{OR}_{TY} \). Now, however, the observations include \( T \); the problem is that they are confined to the \( X = 0 \) stratum, so that the observed counts are \( A_{101}, A_{100}, A_{001}, A_{000} \).

With no data at \( X = 1 \), the non-identified parameters from the log-linear model are those involving \( X \) \((\beta_X, \beta_{TX}, \beta_{XY}, \beta_{TXY})\), which are the coefficients in the logistic regression of non-response on \( T \) and \( Y \),

\[
\Pr(X = 1 | T = t, Y = y) = \expit(\beta_X t + \beta_{TX} t y + \beta_{XY} y + \beta_{TXY} ty).
\]

Bayesian analysis may proceed by fitting the full log-linear model (1) to the observed \( TXY \) data. The actual records are now complete, but all have \( X = 0 \). There will also be added records representing priors for \( \beta_X, \beta_{TX}, \beta_{XY}, \beta_{TXY} \) and perhaps other coefficients. For independent priors, each prior record will have all covariate entries set to zero except the covariate whose coefficient prior it represents.

The \( TY \) odds ratios among respondents and non-respondents are \( \text{OR}_{T00} = \exp(\beta_{TY}) \) and \( \text{OR}_{T11} = \exp(\beta_{TX} + \beta_{TXY}) = \exp(\beta_{TY}) \exp(\beta_{XY}) \). If the proportion sampled at each \( TY \) level is small, the distribution among the non-respondents will approximate that of the population. In particular, the non-respondent odds ratio \( \exp(\beta_{TY} + \beta_{TXY}) \) will approximate the marginal (population) \( \text{OR}_{TY} \). In that case the respondent odds ratio \( \exp(\beta_{TY}) \) has approximate bias \( \exp(-\beta_{TXY}) \).

As a consequence, if only \( \text{OR}_{TY} \) is of interest, coefficient priors are independent, and no prior is used on the identified coefficients, the other non-identified coefficients \((\beta_X, \beta_{TX}, \beta_{XY})\) can be dropped from the model without changing inferences on \( \text{OR}_{TY} \).\(^7\)
Information on non-respondents (subjects with \( X = 1 \)) is sometimes obtained from general records or from call-back surveys of non-respondents. Nonetheless, respondents in such surveys are unlikely to be a random sample of all the original non-respondents; hence a further model will be needed to relate survey non-response to \( T, X \) and \( Y \), as in non-ignorable non-response models.\(^{46, 54-56}\)

**Further topics**

**Multiple biases and multiple variables**

Within the missing-data framework for bias modelling, the covariates, outcomes and measurements may be vectors with arbitrary components. For example, each measurement \( Z \) (whether an exposure, confounder or outcome) with modelled errors has a corresponding true value \( T_Z \). This true variable will have a missing code in actual data, and will have an imputation model with non-identified parameters. Validation merely adds records with \( T_Z \) or another measure of it. In a parallel fashion, each modelled but unmeasured confounder will have its own column with a missing code in actual data, and will have its own imputation model with non-identified parameters; second-phase data merely add records with the confounder. Likewise, the model for non-response or selection will be non-identified and may incorporate dependencies on covariates, outcomes and their measurements.

**Sporadically missing data**

Actual-data records may include among their missing values some that are sporadically missing. These sporadic missing values can be handled as part of the same analysis, although the computing burden to allow for them may be large.\(^{45, 46}\)

**Computational accuracy**

In the above example, all techniques gave similar answers because of the normal priors and large observed numbers. Nonetheless, in examples involving asymmetric priors or sparse data sets, normal approximations to coefficient posterior distributions may become unacceptably inaccurate. In those situations one can employ accelerated bootstrap estimation\(^{57}\) on the augmented data (re-estimating the coefficients at each resampling) or penalized profile-likelihood intervals;\(^{41}\) results from these more accurate approximations also provide checks and starting values for direct posterior samplers. Conversely, direct posterior sampling provides checks on approximate methods. Such as cross-checks can be especially important if the posterior distribution may be multimodal.

All methods require attention to convergence as well as accuracy. For example, convergence criteria for iterative analytic methods should be set very tightly to avoid premature termination.\(^{58}\) The potential for failure of any method (including Markov chain as well as analytic methods) suggests that computing results in different ways is advisable outside of the simplest problems.

**Discussion**

I believe that all analyses should be viewed as parts of a sensitivity analysis,\(^{59}\) in recognition of the fact that we cannot know our assumptions are correct. Without identification we cannot even tell from our data whether our assumptions are reasonable approximations. At best we can only propose models that incorporate or are at least consistent with the facts as we know them.

Unfortunately, there will be an infinite number of these models and they will not all be mutually compatible. As a consequence, statistics provides only inferential possibilities rather than inferences. Bias analysis permits formal (which is to say, explicit, public and deductive) possible inferences from models that are not grossly overconfident or absurd. This advantage becomes most important when the confidence intervals from conventional analysis are narrow or \( P \)-values are small (as is common in large studies or meta-analyses), so that biases rather than random errors become the dominant source of uncertainty.

Multiple-bias models often contain more parameters than observed counts, as above. This complexity again reflects that there are many more sources of uncertainty than can be accounted for by the data alone. Uncertainty can often be adequately addressed, however, by rather simple analyses with one or a few biases. As above, those analyses may quickly reveal that the data cannot sustain any accurate inference about the target parameter given reasonable uncertainties about the bias sources.

The non-identification of bias parameters precludes use of so-called objective methods such as ordinary frequentist methods or ‘objective’ Bayesian statistics (which are based on non-informative priors or ‘reference’ priors). It is neither objective nor scientific to ignore the wealth of background information in forming inferences when the only alternatives are to either discard the data or rely on misleading conventions in order to use the data. Above, a non-informative prior for the predictive parameters \( \pi_{xy} \) produces a 95% interval for \( OR_{xy} \) of \((0, \infty)\), effectively discarding the data, whereas a conventional analysis (which corresponds to using point priors on the \( \pi_{xy} \)) produces the grossly overconfident interval of \((1.1, 1.8)\).

Only by judicious use of well-informed priors between these extremes can we obtain well-informed inferences.

In summary, a major problem with conventional frequentist and Bayesian analyses is that they conceal dogmatic point-null priors on hidden bias parameters.
Bias analysis is not limited to such overoptimistic extremes. It thus frees researchers from having to use the ludicrous priors implicit in conventional results, and shows how classic validity problems can be subsumed under the topic of analysis with missing data. I thus advocate that bias modelling be covered in basic statistical training in the health sciences, and that epidemiologists understand principles of bias analysis so that they fully appreciate the complexity of inference from observational data.

Nonetheless, I do not believe that every study should perform a bias analysis. Bias analysis is superfluous when conventional intervals show that no useful conclusion could be drawn from the study even it were perfect apart from random error. More generally, rather than providing a bias analysis, a study may be of greater service by refraining from inference; instead it can focus on carefully reporting its design, conduct and data in great detail to facilitate pooling and meta-analysis. Inferences are best based on a more complete account of evidence than can be provided in a single study report. Although illustration is simplest within a single study, the effort of bias analysis is more justifiable in research synthesis. Even there, bias analysis becomes essential only when doing risk assessment or when authors claim to offer definitive conclusions regarding the target of study.

Acknowledgements

The author wishes to thank Steve Cole, Michael Höfler, Katherine Hoggatt and the reviewers for their helpful comments.

Conflict of interest: None declared.

References

Appendix 1

General data priors

The data-prior approach is not limited to conjugate priors. In particular, the likelihood from prior data need not be from the same distributional family as the likelihood from the actual data; the two

references:

likelihoods need not even share all parameters.\textsuperscript{10} Any prior density that can be represented as a likelihood function from a statistical experiment can be converted into prior data and appended to the original data set. Frequency coverage under the implied frequentist model (treating the prior as a likelihood contribution) will then become approximate posterior coverage. Furthermore, conjugate priors can be generalized with added parameters and still retain a data representation.\textsuperscript{41,42}

As an example, the generalized-conjugate (logit-beta) prior \( p(\beta) \) for binomial data is

\[
p(\beta; m, s, n, t) \propto \left\{ \frac{\text{expit}(t)(\beta^{s+t})}{1 + \text{expit}(t)} \right\}^n = \frac{e^{H(\beta/s+t)}}{(1 + e^{H(\beta)+t})^s},
\]

where \( \beta^* \equiv (\beta - m)/s \), \( A \equiv n \times \text{expit}(t) = n e^t/(1 + e^t) \) and \( H \equiv (t - m)/s \). The parameter \( m \) is the mode; \( s \) controls the scale or spread; \( n \) controls tail-weight, with lighter tails for larger \( n \); and \( t = \logit(A/n) \) controls skewness, with \( t = 0 \) for symmetry, \( t < 0 \) for right-skewed and \( t > 0 \) for left-skewed.\textsuperscript{41} This distribution is derivable from \( \beta = m + s \times \ln(\phi) \) where \( \phi \) is an \( F \) variate with 2\( A \), 2\( (n - A) \) degrees of freedom.\textsuperscript{41} As a special case, \( p(\beta; m, s, n, t) \) with \( s = 1, n = 2, t = 0 \) is the standard logistic distribution, for which \( \beta \sim \logit(\theta) \) where \( \theta \sim \text{uniform}(0, 1) \). On the other hand, the distribution rapidly approaches normality with mean \( \mu = m \) and variance \( s^2/A(1 - A/n) \) as \( n \) increases with \( t \) fixed; e.g., with \( t = 0 \) and \( n = 10 \), it matches to 2 digits the 2.5th and 97.5th percentiles of a normal with the same mean and variance.

Suppose we observed \( A \) successes from a binomial distribution with \( n \) trials and success probability \( e^{0s + H}/(1 + e^{0s + H}) \). The likelihood from this observation is then \( L(\beta) = e^{H(\beta/s+t)}/(1 + e^{H(\beta)+t})^s \), which is proportional to the above prior density. Maximization of \( L(\beta) \) yields a mode at \( m \) with inverse information \( \tau^2 = s^2/\ln(A/n) \). Hence a binomial logistic-regression program will add the log prior \( \ln\{p(\beta; m, s, n, t)\} \) to the actual-data loglikelihood if it is given a prior-data record having \( A \) successes out of \( n \) total observed, an offset of \( H = (t - m)/s \), with the covariate of coefficient \( \beta \) set to 1/s and with all other covariate values in the record set to zero (including the ‘constant’ \( D \)). If \( A \) or \( n \) is not an integer then one must use software that does not truncate input counts.

The offset \( H \) is a variable with coefficient \( \gamma \) constrained to equal one (p. 206).\textsuperscript{52} We can impose \( \gamma = 1 \) by adding a binomial prior record for \( \gamma \) with \( 10^6 \) ‘successes’, \( 10^6 \) ‘failures’, \( H = 1 \), and all other covariates (including \( D \)) equal to zero.\textsuperscript{41,42} \( H \) is unnecessary if it is zero for all priors. For symmetric \( (t = 0) \) priors, \( H = -m/s \); thus \( H \) is unnecessary if all priors are mode-zero and symmetric.

The binomial likelihood for \( \beta \) from a prior observation \( A \) on \( n \) trials will be almost normal(\( \mu, \tau^2 \)) if \( t = 0\) \( H = -\mu/s \) and \( 2A = n = 4s^2/\tau^2 \geq 800 \); hence \( s \geq 10 \) will suffice for \( \tau^2 \leq 0.5 \). Similarly, for Poisson data the likelihood for \( \beta \) from a prior observation \( A \) will be almost normal(\( \mu, \tau^2 \)) if \( H = \ln(A) - \mu/s \) and \( A = s^2/\tau^2 \geq 1600 \); hence \( s \geq 30 \) will suffice for \( \tau^2 \leq 0.5 \). If a time-on-test (person-time) variable is called for by the program, we can dispense with entering \( H \) separately and use instead a time of \( e^H = A/\exp(m/s) \) for the prior records.

To use the more general prior \( p(\beta; m, s, n, t) \) for Poisson regression we must combine Poisson and binomial likelihoods. A product of Poisson likelihoods can be represented as a likelihood with inverse information \( \tau^2 = s^2/\ln(n) \). Thus a binomial record for \( \gamma \) of coefficient \( \gamma \) with \( 10^6 \) ‘successes’ and \( 10^6 \) ‘failures’ will be almost normal(\( \mu, \tau^2 \)) if \( H = -\mu/s \) and \( 2A = n = 4s^2/\tau^2 \geq 800 \); hence \( s \geq 10 \) will suffice for \( \tau^2 \leq 0.5 \). If a time-on-test (person-time) variable is called for by the program, we can dispense with entering \( H \) separately and use instead a time of \( e^H = A/\exp(m/s) \) for the prior records.