RT-MAE 2010-04

INFERENTIAL IMPLICATIONS OF OVER-PARAMETERIZATION: A CASE STUDY IN INCOMPLETE CATEGORICAL DATA

by

Frederico Z. Poleto, Carlos Daniel Paulino, Geert Molenberghs
and
Julio M. Singer

Palavras-Chave: Contingency table, Identifiability, Incomplete data, Pattern-mixture model, Selection model, Sensitivity analysis.

Classificação AMS: 62F12, 62F15.

- Junho de 2010 -

Inferential implications of over-parameterization: a case study in incomplete categorical data

Frederico Z. Poleto[1], Carlos Daniel Paulino[2],
Geert Molenberghs[3] and Julio M. Singer[1]

[1]Instituto de Matemática e Estatística, Universidade de São Paulo,
Caixa Postal 66281, São Paulo, SP, 05314-970, Brazil

[2]Instituto Superior Técnico, Universidade Técnica de Lisboa (and CEAUL-FCUL),
Av. Rovisco Pais, Lisboa, 1049-001, Portugal

[3]I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium,
and Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

Abstract

In the context of either Bayesian or classical sensitivity analyses of over-parametrized models for incomplete categorical data, it is well known that prior-dependence on posterior inferences of non-identifiable parameters or that too parsimonious over-parameterized models may lead to erroneous conclusions. Nevertheless, some authors either pay no attention to which parameters are non-identifiable or do not appropriately account for possible prior-dependence. We review the literature on this topic and consider simple examples to emphasize that in both inferential frameworks, the subjective components can influence results in non-trivial ways, irrespectively of the sample size. Specifically, we show that prior distributions commonly regarded as slightly informative or non-informative may actually be too informative for non-identifiable parameters, and that the choice of over-parameterized models may drastically impact the results, suggesting that a careful examination of their effects should be considered before drawing conclusions.

Key words: Contingency table, Identifiability, Incomplete data, Pattern-mixture model, Selection model, Sensitivity analysis.

1 Introduction

Models that take the incomplete data generating process into account are, in their most general form, over-parameterized and non-identifiable. Paulino & Pereira (1994) present a nice review on non-identifiability in Statistics and discuss its consequences in both classical and Bayesian analyses. Neath & Samaniego (1997) and Gustafson (2005) also present an interesting discussion centred on the Bayesian point of view, while Daniels & Hogan (2007) specialize focus on longitudinal missing data. For drawing classical inferences, the most common way to overcome non-identifiability is to consider identifying constraints that allow the model to reflect some specific missing data generating mechanism. Because the underlying assumptions are generally unverifiable, statisticians usually perform an "informal" sensitivity analysis based on a set of plausible (but subjective) identifiable models, even though such models may still fail to reflect some desired and more complex missingness mechanism. A more formal alternative, which we call *classical sensitivity analysis*, involves repeated sensitivity analyses with different over-parameterized models for which the non-identifiable parameters are set to fixed values (see e.g., Nordheim, 1984; Copas & Li, 1997; Scharfstein et al., 1999; Vansteelandt et al., 2006). These routes may appear to be unnecessary under a Bayesian framework because the use of proper prior distributions unblocks the inferential process (Paulino & Pereira, 1995). However, given that the data do not contain information to update the prior distribution for all parameters, additional care must be taken in its elicitation. Indeed, non-identifiability may be hidden in the proposed priors and this may create a false sense of precision if very careful attention is not given to their choice.

Molenberghs et al. (2001) state that more over-parameterization yields more uncertainty, whereas too parsimonious models may miss the true one. In practice, however, these and other authors (e.g., Daniels & Hogan, 2007, Chap. 10) usually propose a reduction in the dimension of the adopted models and consider only a few parameters for their sensitivity analyses. Keeping this in mind, our first objective is to illustrate further that such reduction of the dimension must be carried out with great caution to avoid misleading conclusions. Although we scrutinize this only under a classical approach, similar consequences are expected under a Bayesian framework, too.

When considering Bayesian analyses of incomplete categorical data, Paulino & Pereira (1992, 1995) and Forster & Smith (1998) clearly indicate that posterior summaries of parameters of interest are prior-dependent. Soares & Paulino (2001) also note this, but erroneously believe that the implications of this dependence may be mild for large samples, given that the data update some functions of the parameters of interest. Indeed, Neath & Samaniego (1997) conclude:

> *Bayesian analysis cannot be used with impunity in estimating non-identifiable parameters. ... Because posterior estimates of non-identifiable parameters are strongly influenced by prior modelling, even as the sample size grows without bound, it is important to use the utmost care in applying and interpreting Bayesian analysis in such settings.*

Recently, Tian et al. (2003) and Jiang & Dickey (2008) consider higher levels of over-parameterization (allowing for misclassification as well as for missingness in categorical data) without clarifying that the parameters of interest are non-identifiable and without appropriately accounting for the prior-dependence. Our second objective is to show that a good deal of prior-dependence still remains for non-identifiable parameters in analyses with huge sample sizes and priors commonly considered as non-informative or slightly informative. We do this by analyzing simple examples with missing categorical responses, although the essence of our conclusions will carry over to missing continuous responses, as well as to other statistical areas where non-identifiability is of concern, like those where misclassification and measurement errors are involved.

We review both the Bayesian and the classical approaches for the analysis of incomplete categorical data in Section 2. We explore and compare both frameworks using illustrative data in Section 3 and reanalysing the data from the Collaborative Perinatal Project displayed in Table 1 (Baker et al., 1992) in Section 4. In the first case, we examine the effects of a monotone missingness pattern on identifiable and non-identifiable parameters with data from samples with sizes varying from small to large. In the second case, a prospective study with pregnant women to assess the association between maternal smoking and low newborn weight, we illustrate that the conclusions may change depending on the choice of the subjective components of the Bayesian and the classical approaches.

Table 1: Observed frequencies from the Collaborative Perinatal Project.

Smoker	Newborn weight (kg)		
mother	< 2.5	≥ 2.5	missing
yes	4,512	21,009	1,049
no	3,394	24,132	1,135
missing	142	464	1,224

2 Inferential Approaches

Paulino & Pereira (1992) developed a Bayesian solution based upon unconstrained missingness models for incomplete categorical data analysis when the missingness patterns can be structured into partitions of the set of categories. Examples of these patterns are the ones generated by incomplete classification only into marginal subtables as in Table 1. Such a requirement was relaxed in Paulino & Pereira (1995), who considered more general censoring patterns. Their approach serves as the basis of Section 2.1, where we sketch the problem, define notation, and specify the likelihood; it also serves as a starting point for the identification of the prior and posterior distributions discussed in Section 2.2. In Section 2.3, we review the classical sensitivity analysis of Vansteelandt et al. (2006). Finally, in Section 2.4, we comment on differences between both approaches and on extensions.

2.1 Problem description, notation, and likelihood

Consider a random sample of size n, wherein each of the units is classified into response category r with probability θ_r, $r = 1, \ldots, R$, and R corresponds to the number of combinations of the levels of the response variables. For several reasons attributable to censoring or missingness mechanisms, we may only be able to observe the frequency of units in nonempty subsets C of $\{1, \ldots, R\}$, which we call response classes. In particular, the response for a unit is fully categorized (in category r) or fully missing depending on whether $C = \{r\}$ or $C = \{1, \ldots, R\}$. We assume that units with response in category r are observed in class C with probability $\lambda_{C(r)}$. We also assume that there is no misclassification, i.e., $\lambda_{C(r)} = 0$ whenever $r \notin C$. Taking $\mathcal{P}_o$ as the union of the response classes with no missingness and $\mathcal{P}_m$ as the union of those with some degree of missingness, it follows that $\mathcal{P} = \mathcal{P}_o \cup \mathcal{P}_m$ embraces all possible response patterns. Likewise, the data may be summarized in the vector $\mathrm{N} = (\mathrm{N}_o', \mathrm{N}_m')'$, where $\mathrm{N}_o = (n_C, C \in \mathcal{P}_o)' = (n_r, r = 1, \ldots, R)'$ stacks the frequencies of the fully categorized observations, and $\mathrm{N}_m = (n_C, C \in \mathcal{P}_m)'$ encloses the frequencies of the partially or completely missing observations. Let $\boldsymbol{\theta} = (\theta_r, r = 1, \ldots, R)'$ be the parameter of interest, and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_r', r = 1, \ldots, R)'$ be the vector of conditional probabilities of missingness, where $\boldsymbol{\lambda}_r = (\lambda_{C(r)}, C \in \mathcal{P}_r)'$, and $\mathcal{P}_r = \{C \in \mathcal{P} : r \in C\}$ contains the response classes that include category r. Note that the natural constraints are $\sum_{r=1}^{R} \theta_r = 1$ and $\sum_{C \in \mathcal{P}_r} \lambda_{C(r)} = 1$, $r = 1, \ldots, R$. Hence, the likelihood function for $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is

$$L(\boldsymbol{\theta}, \boldsymbol{\lambda} | \mathrm{N}) \propto \prod_{C \in \mathcal{P}} \left(\sum_{r \in C} \theta_r \lambda_{C(r)} \right)^{n_C} = \prod_{r=1}^{R} (\theta_r \lambda_{r(r)})^{n_r} \times \prod_{C \in \mathcal{P}_m} \left(\sum_{r \in C} \theta_r \lambda_{C(r)} \right)^{n_C}.$$

This parameterization is known as selection model (Glynn et al., 1986; Little & Rubin, 2002). Alternatively, we may consider the joint probabilities $\boldsymbol{\mu} = \{\mu_{Cr}\}$, where $\mu_{Cr} = \theta_r \lambda_{C(r)}$, or yet a version of the pattern-mixture model parameterization (Paulino & Pereira, 1995) that is suitable to identify parametric functions for which the sample contains no information.

Let $\gamma_o = \sum_{r=1}^{R} \mu_{rr}$ be the probability that a unit is fully categorized, and $\gamma_C = \sum_{r \in C} \mu_{Cr}$, be the probability of observing the response class $C, C \in \mathcal{P}_m$. Likewise, let $\alpha_o = (\alpha_{r(o)}, r = 1, \ldots, R)'$, with $\alpha_{r(o)} = \mu_{rr}/\gamma_o$, be the conditional probabilities of response categories $r = 1, \ldots, R$ given a complete classification, and $\alpha_C = (\alpha_{r(C)}, r \in C)'$, with $\alpha_{r(C)} = \mu_{Cr}/\gamma_C$, denote the conditional probabilities of belonging to each of the categories included in the response class C given this partial categorization, $C \in \mathcal{P}_m$. Then, the likelihood function of (γ, α), where $\gamma = (\gamma_o, \gamma_C, C \in \mathcal{P}_m)'$ and $\alpha = (\alpha_o', \alpha_C', C \in \mathcal{P}_m)'$, is expressed as

$$L(\gamma, \alpha | \mathrm{N}) \propto \gamma_o^{n_o} \prod_{C \in \mathcal{P}_m} \gamma_C^{n_C} \times \prod_{r=1}^{R} \alpha_{r(o)}^{n_r} \times \prod_{C \in \mathcal{P}_m} \left(\sum_{r \in C} \alpha_{r(C)} \right)^{n_C}, \qquad (1)$$

where $n_o = \sum_{r=1}^{R} n_r$. Since the natural constraints are $\gamma_o + \sum_{C \in \mathcal{P}_m} \gamma_C = 1$, $\sum_{r=1}^{R} \alpha_{r(o)} = 1$, and $\sum_{r \in C} \alpha_{r(C)} = 1$, $C \in \mathcal{P}_m$, it is clear that the parameters $\{\alpha_{r(C)}, C \in \mathcal{P}_m\}$ actually do not appear in (1) and, hence, there is no sample information about them.

To illustrate the concepts introduced so far, we turn to the data presented in Table 1, where we replace the index r by two indices, namely, $i = 1, 2$, to indicate smoking and non-smoking mothers, respectively, and $j = 1, 2$, to indicate birthweight < 2.5 or ≥ 2.5 kg, respectively. Then, the response classes associated to the data in Table 1 are displayed in Table 2, so that, $\mathbf{N}_o = (4,512; 21,009; 3,394; 24,132)'$ and $\mathbf{N}_m = (1,049; 1,135; 142; 464; 1,224)'$ are the frequencies corresponding to the response classes in $\mathcal{P}_o = \{\{11\}, \{12\}, \{21\}, \{22\}\}$ and $\mathcal{P}_m = \{\{11, 12\}, \{21, 22\}, \{11, 21\}, \{12, 22\}, \{11, 12, 21, 22\}\}$, respectively.

Table 2: Response classes from the Collaborative Perinatal Project.

Smoker mother	Newborn weight (kg) < 2.5	≥ 2.5	missing
yes	{11}	{12}	{11, 12}
no	{21}	{22}	{21, 22}
missing	{11, 21}	{12, 22}	{11, 12, 21, 22}

2.2 Prior and posterior distributions

Paulino & Pereira (1995) emphasized that assuming a Dirichlet prior distribution with hyper-parameters $\{a_{Cr}\}$ for $\boldsymbol{\mu}$, i.e.,

$$\text{Prior 1}: \boldsymbol{\mu} \sim D(\{a_{Cr}\}),$$

is equivalent to setting

$$\text{Prior 1}: \begin{cases} \boldsymbol{\theta} \sim D(a_{+r}, r = 1, \ldots, R), \text{ where } a_{+r} = \sum_{C \in \mathcal{P}_r} a_{Cr}, \\ \boldsymbol{\lambda}_r \sim D(a_{Cr}, C \in \mathcal{P}_r), \; r = 1, \ldots, R, \end{cases}$$

all mutually independent, and also to assuming

$$\text{Prior 1}: \begin{cases} \boldsymbol{\gamma} \sim D(a_o, a_{C+}, C \in \mathcal{P}_m), \text{ where } a_o = \sum_{r=1}^{R} a_{rr} \text{ and } a_{C+} = \sum_{r \in C} a_{Cr}, \\ \boldsymbol{\alpha}_o \sim D(a_{rr}, r = 1, \ldots, R), \\ \boldsymbol{\alpha}_C \sim D(a_{Cr}, r \in C), \; C \in \mathcal{P}_m, \end{cases}$$

all mutually independent. The posterior distribution of $\boldsymbol{\mu}$, and also that of $\boldsymbol{\theta}$, can be expressed as a generalized Dirichlet distribution (i.e., a finite mixture of Dirichlet distributions) as defined by Dickey (1983) and has no simple closed form; Tian et al. (2003) took this approach further. The pattern-mixture model parameterization leads to $\boldsymbol{\gamma}|\mathrm{N} \sim D(a_o + n_o, a_{C+} + n_C, C \in \mathcal{P}_m)$, $\boldsymbol{\alpha}_o|\mathrm{N} \sim D(a_{rr} + n_r, r = 1, \ldots, R)$, $\boldsymbol{\alpha}_C|\mathrm{N} \sim D(a_{Cr}, r \in C)$, $C \in \mathcal{P}_m$, all mutually independent. Paulino & Pereira (1992, 1995) used the relation $\theta_r = \gamma_o \, \alpha_{r(o)} + \sum_{C \in \mathcal{P}_r \cap \mathcal{P}_m} \gamma_C \, \alpha_{r(C)}$ to obtain the mean and the covariance matrix of $\boldsymbol{\theta}|\mathrm{N}$. Using also $\lambda_{r(r)} = \gamma_o \alpha_{r(o)}/\theta_r$ and $\lambda_{C(r)} = \gamma_C \alpha_{r(C)}/\theta_r$, $C \in \mathcal{P}_r$, Soares & Paulino (2001) adopted a Monte Carlo simulation approach that easily allows one to draw inferences on general functions of $(\boldsymbol{\theta}, \boldsymbol{\lambda})$. Even knowing that the prior distribution of $\{\boldsymbol{\alpha}_C, C \in \mathcal{P}_m\}$ is not updated by the data, they believed that the possible prior-dependence implications on the posterior distribution of (θ, λ) should be mild for large samples, given the actual updating of γ and α_o.

The expert is not free to make individual judgments about the most meaningful parameters (θ and λ) with Prior 1 and the existing connection among the hyper-parameters is not generally justifiable from a practical point of view as discussed by Soares & Paulino (2007) and Jiang & Dickey (2008). For this reason, they adopted

$$\text{Prior 2}: \begin{cases} \theta \sim D(b_r, r = 1, \ldots, R), \\ \lambda_r \sim D(a_{Cr}, C \in \mathcal{P}_r), \ r = 1, \ldots, R, \end{cases}$$

all mutually independent; these authors and Tian et al. (2003) also allowed for misclassification in their modelling strategy. Jiang & Dickey (2008) worked with the resulting generalized Dirichlet posterior distribution, while Soares & Paulino (2007) used a chained data augmentation algorithm (Tanner & Wong, 1987) to obtain a sample of the posterior distribution, wherein the augmented data are the hypothetical frequencies $\mathrm{m} = \{m_{Cr}\}$ of units with responses in category r and classified in C. Apart from the case when there are no missing data ($m_{rr} = n_r$), these frequencies are non-observable and we only know that $\sum_{r \in C} m_{Cr} = n_C$, $C \in \mathcal{P}_m$. Given (θ, λ) and N, $\{m_{Cr}, r \in C\}$ are multinomially distributed with parameters n_C and $\alpha_C(\theta, \lambda)$, $C \in \mathcal{P}_m$, and $\theta|\mathrm{m} \sim D(b_r + m_{+r}, r = 1, \ldots, R)$, where $m_{+r} = \sum_{C \in \mathcal{P}_r} m_{Cr}$, and $\lambda_r|\mathrm{m} \sim D(a_{Cr} + m_{Cr}, C \in \mathcal{P}_r)$, $r = 1, \ldots, R$, are all mutually independent. Hence, the algorithm proposed by Soares & Paulino (2007) consists of sampling alternately from these distributions. Even though there is no sample information about α_C, its prior and posterior distributions are no longer the same in this case, because γ, α_o, and α_C are no longer mutually independent a priori and lead to indirect learning through the dependence among them, as Scharfstein et al. (2003) pointed out.

Prior 2 seems to be flexible enough to represent information elicited by experts or obtained from historical data. However, when such information cannot be obtained, as in most practical cases, the question is what should be regarded as a non-informative prior distribution. Natural choices may be to set all $\{b_r\}$ and $\{a_{C_r}\}$ equal to 1, 0.5, or some small value, say 0.1. The first choice leads to Bayes-Laplace's prior, i.e., uniform distributions on the corresponding simplices. The second resembles Jeffreys' prior in the complete-data case. The third is even more diffuse, in the direction of Haldane's prior, which cannot be used because non-identifiability does not allow us to consider improper priors for all parameters, setting all $\{b_r\}$ and $\{a_{C_r}\}$ equal to 0. If we turn to Prior 1, the corresponding choices are even less clear because the connections among the hyper-parameters do not allow, for example, to use the same values for the hyper-parameters of θ and $\{\lambda_r\}$.

The problem is that Jeffreys' prior cannot be computed for non-identifiable models, because the Fisher information matrix is singular. In a different context, Wang & Ghosh (2000) overcame the same problem by (i) choosing some non-identifiable parameters to turn the others conditionally identifiable, (ii) specifying a conditional version of Bayes-Laplace's prior to the selected non-identifiable parameters given the conditionally identifiable ones, (iii) computing the marginal likelihood function of the conditionally identifiable parameters, and (iv) obtaining Jeffreys' marginal prior for them. Although there is no unique choice, an option is to consider θ and, for each class $C \in \mathcal{P}_m$, one $\lambda_{C(r)}$, $r \in C$, in the conditionally identifiable parameter set, and the other $\lambda_{C(r)}$, $C \in \mathcal{P}_m$, in the selected non-identifiable parameter set; $\{\lambda_{r(r)}\}$ may be obtained using the natural constraints, and therefore they are also conditionally identifiable. Given the conditionally identifiable parameters, the Bayes-Laplace prior for the selected non-identifiable parameters is a uniform distribution on the space obtained by the direct product of the "reduced" simplices of λ_r, $r = 1, \ldots, R$, generated by fixing the conditionally identifiable parameters. Although this route could be taken further, it is simpler to switch to the pattern-mixture model parameterization if one believes that the independence between (γ, α_o) and $\{\alpha_C, C \in \mathcal{P}_m\}$ is reasonable. Because the parameter spaces of (γ, α_o) and $\{\alpha_C, C \in \mathcal{P}_m\}$ are not related, the independence assumption between them leads to the same marginal likelihood (1) under any prior distribution for $\{\alpha_C, C \in \mathcal{P}_m\}$. Jeffreys' prior is then $\gamma \sim D(R/2, \{0.5, C \in \mathcal{P}_m\})$ and $\alpha_o \sim D(0.5, r = 1, \ldots, R)$, considered mutually independent. This fits into Prior 1 with $a_{rr} = 0.5$, $r = 1, \dots, R$, and $\sum_{r \in C} a_{Cr} = 0.5$, $C \in \mathcal{P}_m$. To complete specification of Prior 1, a natural choice is $a_{Cr} = 0.5/\#C$, $C \in \mathcal{P}_m$, where $\#C$ is the cardinality of the class C. We may also break the connections among the hyper-parameters of Prior 1 and consider

$$\text{Prior 3}: \begin{cases} \gamma \sim D(R/2, \{0.5, C \in \mathcal{P}_m\}), \\ \alpha_o \sim D(0.5, r = 1, \dots, R), \\ \alpha_C \sim D(c_{Cr}, r \in C),\ C \in \mathcal{P}_m, \end{cases}$$

all mutually independent. This gives us more flexibility to use Bayes-Laplace's prior ($c_{Cr} = 1$), a Jeffreys analogue ($c_{Cr} = 0.5$) or more diffuse priors (e.g., $c_{Cr} = 0.1$) for these non-identifiable parameters. Scharfstein et al. (2003) presented reasons why the above mentioned independence assumption may not be adequate in practice, but this does not seem to have an impact on analyses with non-informative priors. Like in Prior 1 with this parameterization, posterior distributions are easily obtained, leading to $\gamma|\mathrm{N} \sim D(R/2 + n_o, 0.5 + n_C, C \in \mathcal{P}_m)$, $\alpha_o|\mathrm{N} \sim D(0.5 + n_r, r = 1, \dots, R)$, $\alpha_C|\mathrm{N} \sim D(c_{Cr}, r \in C)$, $C \in \mathcal{P}_m$, all mutually independent.

2.3 Classical sensitivity analysis

To motivate the present discussion, we consider again the data in Table 1 and the notation introduced in the last paragraph of Section 2.1. Under any of the parameterizations considered in that section, there are 15 parameters not functionally related, but only 8 observed frequencies (the total, $n = 57,061$, is fixed). Therefore, any model with 9 or more parameters is over-parameterized. Let us recall the construction of conditional models considered in Fay (1986) where, as suggested by Molenberghs et al. (1999), the elements of $\boldsymbol{\lambda}$ are reparameterized by $\psi_{1(ij)}$, the probability of observing the maternal smoking status, $\psi_{21(ij)}$, the probability of observing the newborn weight given that the maternal smoking status is observed, and $\psi_{20(ij)}$, the probability of observing the newborn weight given that the maternal smoking status is missing. These probabilities are also conditioned on the maternal smoking status and the newborn weight. We consider the following two models:

$$\text{Model 1}: \begin{cases} \text{logit}\left(\psi_{1(ij)}\right) &= \phi_{10} + \phi_1(i-1) + \phi_2(j-1) + \phi_3(i-1)(j-1), \\ \text{logit}\left(\psi_{21(ij)}\right) &= \phi_{20} + \phi_1(i-1) + \phi_2(j-1) + \phi_3(i-1)(j-1), \\ \text{logit}\left(\psi_{20(ij)}\right) &= \phi_{30} + \phi_1(i-1) + \phi_2(j-1) + \phi_3(i-1)(j-1), \end{cases}$$

$$\text{Model 2}: \begin{cases} \text{logit}\left(\psi_{1(ij)}\right) &= \phi_{10} + \phi_{11}(i-1) + \phi_{12}(j-1), \\ \text{logit}\left(\psi_{21(ij)}\right) &= \phi_{20} + \phi_{21}(i-1) + \phi_{22}(j-1), \\ \text{logit}\left(\psi_{20(ij)}\right) &= \phi_{30} + \phi_{21}(i-1) + \phi_{22}(j-1), \end{cases}$$

which have, respectively, 9 and 10 non-redundant parameters when combined with $\boldsymbol{\theta}$.

Kenward et al. (2001) and Molenberghs et al. (2001) distinguished between two types of statistical uncertainty: statistical imprecision and statistical ignorance. The former is caused by not observing the entire population, and is usually quantified by standard errors and confidence regions. The latter is due to deficiencies in the observation process; e.g., when some responses are missing, misclassified, and/or measured with error. When the sample size tends to infinity, the magnitude of statistical imprecision decreases to zero, but statistical ignorance may not change. In our case, statistical ignorance is related to the distribution of the censored response classes over the appropriate response categories. For example, given that $\boldsymbol{\omega} = \phi_3$ in Model 1, or that $\boldsymbol{\omega} = (\phi_{11}, \phi_{22})$ in Model 2 are fixed, we may estimate $\boldsymbol{\theta}$ and the other estimable parameters as a function of $\boldsymbol{\omega}$ and also obtain a $100(1-\alpha)\%$ confidence region for them. After repeating the analysis for a set $\boldsymbol{\Omega}$ of values for the sensitivity parameters, $\boldsymbol{\omega}$, the range of the estimates of the estimable parameters provides an estimate for the so-called ignorance region, and the union of the $100(1-\alpha)\%$ confidence regions, for the so-called $100(1-\alpha)\%$ uncertainty region.

Vansteelandt et al. (2006) provided appropriate definitions of consistency for the ignorance region and of coverage for the uncertainty region. The estimated regions are termed

Honestly Estimated Ignorance Region (HEIR) and Estimated Uncertainty Region (EURO). For a scalar estimable parameter β of interest, the interval of ignorance obtained by setting ω equal to ω_l and ω_u is denoted by $\text{ir}(\beta,\Omega) = [\beta(\omega_l),\beta(\omega_u)] = [\beta_l,\beta_u]$. The HEIR, $\widehat{\text{ir}}(\beta,\Omega) = [\hat{\beta}_l,\hat{\beta}_u]$, is said to be weakly consistent for the true $\beta_0 = \beta(\omega_0)$, if convergence in probability of $\hat{\beta} = \hat{\beta}(\omega)$ to $\beta = \beta(\omega)$ holds for all $\omega \in \Omega$. Of course, there is always the underlying assumption that $\omega_0 \in \Omega$. Uncertainty regions for β_0 with uncertainty level $100(1-\alpha)\%$ were defined in three different ways: (i) strong EUROs cover $\beta(\omega)$ simultaneously for all $\omega \in \Omega$ with at least $100(1-\alpha)\%$ probability, (ii) pointwise EUROs cover $\beta(\omega)$ uniformly over $\omega \in \Omega$ with at least $100(1-\alpha)\%$ probability, and (iii) weak EUROs have an expected overlap with $\text{ir}(\beta,\Omega)$ of at least $100(1-\alpha)\%$. The authors also provided algorithms for constructing these EUROs, all with the usual form $[\hat{\beta}_l - c_{\alpha^*/2}\,\widehat{\text{se}}(\hat{\beta}_l), \hat{\beta}_u + c_{\alpha^*/2}\,\widehat{\text{se}}(\hat{\beta}_u)]$, where $\hat{\beta}_l$ and $\hat{\beta}_u$ are obtained from consistent and asymptotically normal estimators of β_l and β_u, and $\widehat{\text{se}}(\hat{\beta}_l)$ and $\widehat{\text{se}}(\hat{\beta}_u)$ are obtained from consistent estimators of the standard errors of $\hat{\beta}_l$ and $\hat{\beta}_u$. For strong EUROs, the critical value $c_{\alpha^*/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. This type of interval was used by Kenward et al. (2001) and Molenberghs et al. (2001), among others. For pointwise EUROs, $c_{\alpha^*/2}$ is the solution of

$$\min\left[\Phi(c_{\alpha^*/2}) - \Phi\left(-c_{\alpha^*/2} - \frac{\hat{\beta}_u - \hat{\beta}_l}{\widehat{\text{se}}(\hat{\beta}_u)}\right), \Phi\left(c_{\alpha^*/2} + \frac{\hat{\beta}_u - \hat{\beta}_l}{\widehat{\text{se}}(\hat{\beta}_l)}\right) - \Phi(-c_{\alpha^*/2})\right] = 1-\alpha,$$

where Φ denotes the standard normal cumulative distribution function. For weak EUROs, $c_{\alpha^*/2}$ is the solution of

$$\alpha = \frac{\widehat{\text{se}}(\hat{\beta}_l) + \widehat{\text{se}}(\hat{\beta}_u)}{\hat{\beta}_u - \hat{\beta}_l} \int_0^{+\infty} z\,\varphi(z + c_{\alpha^*/2})\,dz + \varepsilon,$$

where φ is the standard normal density function and ε is the correction term

$$\varepsilon = \int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\text{se}}(\hat{\beta}_u)}^{+\infty} \varphi(z + c_{\alpha^*/2})\, dz - \frac{\widehat{\text{se}}(\hat{\beta}_u)}{\hat{\beta}_u - \hat{\beta}_l} \int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\text{se}}(\hat{\beta}_u)}^{+\infty} z\, \varphi(z + c_{\alpha^*/2})\, dz +$$
$$\int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\text{se}}(\hat{\beta}_l)}^{+\infty} \varphi(z + c_{\alpha^*/2})\, dz - \frac{\widehat{\text{se}}(\hat{\beta}_l)}{\hat{\beta}_u - \hat{\beta}_l} \int_{(\hat{\beta}_u-\hat{\beta}_l)/\widehat{\text{se}}(\hat{\beta}_l)}^{+\infty} z\, \varphi(z + c_{\alpha^*/2})\, dz$$

that may be set equal to zero unless the sample size is small and/or there is little ignorance about β. Strong EUROs are conservative pointwise EUROs, which in turn are conservative weak EUROs. When there is much ignorance about β and the sample size is large, pointwise EUROs approach strong EUROs. The choice among the three versions of EUROs depends on which is the more appropriate definition for the uncertainty region or on the desired degree of conservativeness.

2.4 Differences between Bayesian and classical approaches and extensions

Many criticisms to Bayesian methods rest essentially on the subjective choice of the prior distribution. Such criticisms may also apply to classical sensitivity analyses, because of the subjective choice of values for the sensitivity parameters and for the over-parameterized model.

The set Ω of possible values for the sensitivity parameters may cover an in-depth grid of: (i) the whole parameter space for ω or, (ii) only a "plausible range" elicited by experts, if ignorance with regard to the missing data generating process is not complete. The latter alternative, considered by Vansteelandt et al. (2006) in a simple setting, may sound like a prior distribution for the sensitivity parameter, but actually it only constrains the parameter space and does not account for any further probabilistic reasoning with respect to the more or less probable values. This effort of cautiously eliminating unnecessary ignorance is welcome, but will not be extensively explored here. This can also be done via prior distributions in the Bayesian approach, and, like in this case, can only be fulfilled for parameters that have meaningful interpretations to the experts, which probably would exclude Models 1 and 2 described in Section 2.3.

Models 1 and 2 have got only one and two real-valued sensitivity parameters, respectively. This simplifies the problem from a computational point of view, but has the drawback of restricting attention to these particular families of arbitrary models. Depending on the functions of the parameters to be investigated, the resulting sensitivity analysis may be affected in unforeseen ways as we will show in the next sections. Therefore, unless these models also translate expert beliefs, it may be more reasonable to use the unconstrained (or non-parametric) missingness models of Section 2.1. To accomplish this, we use the strategy described in the last paragraph of Section 2.2, i.e., to split $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ into conditionally identifiable and selected non-identifiable parameters to be respectively employed as estimable and sensitivity parameters; in addition, we may use $\{\log(\lambda_{C(r)}/\lambda_{r(r)}), C \in \mathcal{P}_r \cap \mathcal{P}_m\}$ instead of $\boldsymbol{\lambda}_r$ to easily incorporate the natural constraints and also simplify the task of working with $\boldsymbol{\Omega}$ in a Euclidean space instead of in simplices. An alternative is to use $(\boldsymbol{\gamma}, \boldsymbol{\alpha}_o)$ as estimable parameters and $\{\boldsymbol{\alpha}_C, C \in \mathcal{P}_m\}$ as sensitivity parameters. The corresponding ignorance region, also called non-parametric or best-worst case bounds in such a context, is expectedly much wider, but is a sensible starting point for careful modelling, as Kenward et al. (2001) advocated.

Constrained models, over-parameterized or not, may also be considered under a Bayesian approach, but the prior distributions and/or algorithms described in Section 2.2 need modification. For example, to fit log-linear models to $\boldsymbol{\theta}$, following the proposal of Bedrick et al. (1996), Soares & Paulino (2007) used the induced priors from Prior 2 for the parameters of the log-linear model and then conditioned on the parameters assumed to be null. The slice sampling ideas of Neal (2003) were employed in one of the steps of the chained data augmentation algorithm to sample from the posterior distribution of the log-linear parameters. Slice sampling was also used by Soares (2004) to sample from the posterior distribution of the parameters of constrained missingness models (e.g., missing at random model: $\lambda_{C(r)} = \tau_{C(C)}, \forall r \in \mathcal{C}$). To overcome the hard task of working with priors that account for both the natural and the model constraints, he used ideas of Walker (1996) and penalized the prior distribution of the unconstrained model in the direction of the postulated model.

In summary, both approaches have some degree of subjectivity, but this is an unavoidable characteristic of incomplete data, which we ought to carefully control, as pointed out by Molenberghs et al. (2001). Adopting a particular sensitivity analysis route may be a matter of ease of implementation or adequacy in the context of the problem, or yet of some personal preference. For example, Scharfstein et al. (2003) were dissatisfied for not obtaining a single summary with the ignorance region of the classical strategy, while Vansteelandt et al. (2006) did not want to average out the extremes of the uncertainty interval with Bayesian machinery. In subsequent sections, we will show that the posterior means may easily shift with small changes of the prior distributions, even for large sample sizes. Therefore, we believe that the classical ignorance region is an interesting alternative. Nevertheless, the conclusions obtained under a Bayesian approach shall not be taken from means, but rather from credible intervals or regions, which reflect both statistical ignorance and imprecision. Using a constrained over-parameterized model and/or limiting the range of Ω may obfuscate extremes even more than Bayesian averaging, but such an effect can be avoided in all cases: either using an unconstrained over-parameterized model, or trying to cover the whole parameter space of $\boldsymbol{\omega}$ or using diffuse prior distributions.

3 Illustrative Study

3.1 Description of the analyses

To explore and compare the Bayesian and classical approaches, we consider simulated data sets of sizes n = 40; 400; 4,000; 40,000, and 4,000,000, all with the same relative frequencies exhibited in Table 3.

Table 3: Relative frequencies of simulated data.

$Y_1 \setminus Y_2$	1	2	missing
1	0.05	0.35	0.05
2	0.05	0.40	0.10

Again, we replace the index r by two indices, i, j, to indicate, respectively, the possible values of Y_1 and Y_2, which may be equal to 1 or 2. In this monotone missingness pattern setting, as Y_1 is always observed, its marginal probabilities (e.g., θ_{2+}) are identifiable. However, the difference between marginal probabilities (e.g., $\theta_{+2} - \theta_{2+}$), the odds ratio ($\text{OR} = \theta_{11}\theta_{22}/[\theta_{12}\theta_{21}]$), or its logarithm, are non-identifiable, because they cannot be expressed as a function of the identifiable parameters γ and α_o only. A comparison of the results for these two sets of parametric functions for increasing sample sizes provides a clear-cut distinction between cases where analyses are only affected by statistical imprecision or, additionally, by statistical ignorance.

Under a Bayesian perspective, we consider the non-informative prior distributions mentioned in Section 2.2, namely, those where all hyper-parameters are set equal to 1.0, 0.5, and 0.1 in Priors 2 and 3, as well as our suggestion to complete Jeffreys' prior specification embedded in Prior 1, which here lead to Prior 3 with $\{c_{Cij} = 0.25\}$. We also tried Prior 3 with $\{c_{Cij} = 0.001\}$. To contemplate slightly informative priors, we consider Prior 2 with the hyper-parameters

$$(b_{11}, b_{12}, b_{21}, b_{22}) = (0.1, 0.3, 0.1, 0.5),$$

$$(a_{\{11\}11}, a_{\{11,12\}11}) = (0.6, 0.4), \qquad (a_{\{12\}12}, a_{\{11,12\}12}) = (0.8, 0.2),$$

$$(a_{\{21\}21}, a_{\{21,22\}21}) = (0.7, 0.3), \qquad (a_{\{22\}22}, a_{\{21,22\}22}) = (0.8, 0.2),$$

as well as another version with all values multiplied by 10. Because the hyper-parameters of each distribution ($\boldsymbol{\theta}$ and $\{\boldsymbol{\lambda}_{ij}\}$) sum up to 1 and 10, we label them Inf1 and Inf10, respectively. First, we combine these hyper-parameters for $\{b_{ij}\}$ with $\{a_{Cij} = 0.5\}$, and then, reversely, we use the $\{a_{Cij}\}$ with $\{b_{ij} = 0.5\}$, so we can compare both cases with $\{b_{ij} = a_{Cij} = 0.5\}$ and assess the change of the priors of θ and λ separately. The resulting prior means for θ_{2+}, $\theta_{+2} - \theta_{2+}$ $(= \theta_{12} - \theta_{21})$, and log OR are, respectively, 0.5, 0.0, and 0.0, for the non-informative priors, and 0.6, 0.2, and > 0.5, for the informative $\{b_{ij}\}$. The posterior summaries for these parameters are presented in Tables 4, 6, and 8. Conventional diagnostics were employed to assess convergence for the analyses based on Prior 2 (Heidelberger & Welch, 1983; Geweke, 1992; Raftery & Lewis, 1992). For all priors and sample sizes, the magnitude of the Monte Carlo errors was smaller than the precision of the figures in the tables. In agreement with Agresti & Min (2005) regarding a possible change of conclusions when analysing 1/OR instead of OR via highest posterior density (HPD) intervals, we consider $100(1-\alpha)\%$ equal-tailed credible intervals (CI).

For the classical sensitivity analysis, we consider three over-parameterized models labelled A, B and C. Let $\psi_{2(ij)}$ $(= \lambda_{\{ij\}(ij)})$ be the conditional probability of observing Y_2 given that $Y_1 = i$ and $Y_2 = j$. Then, reparameterizing λ as

$$\text{logit}\left(\psi_{2(ij)}\right) = \phi_0 + \phi_1(i-1) + \phi_2(j-1) + \phi_3(i-1)(j-1),$$

Models A and B are, respectively, obtained by setting $\phi_3 = 0$ and $\phi_2 = 0$; the corresponding sensitivity parameters are $\omega = \phi_2$ and $\omega - \phi_3$. The unconstrained model, with sensitivity parameters $\omega = (\phi_2, \phi_3)$, is labelled Model C. Although ϕ_2 and ϕ_3 can take any real value, we pragmatically use $\Omega = [-20; 20]$ for Models A and B, and $\Omega = [-20; 20]^2$ for Model C. The grids were constructed with increments of 0.1 and 1, respectively. Enlarging Ω did not change the results which are displayed in Tables 5, 7, and 9. Weak EUROs were not computed for θ_{2+} since there is no ignorance about this parameter. For this identifiable parameter, it is interesting to note that pointwise EUROs coincide with strong EUROs, and correspond to intervals obtained via the usual asymptotic normal distribution, and also that the HEIR reduces to a point estimate. We present the results using the exact version of the weak EUROs (i.e., computing ε), but we compare them to the one where $\varepsilon = 0$ and notice negligible differences (< 0.01 for log OR, and < 0.001 for $\theta_{+2} - \theta_{2+}$) for the resulting EUROs under all models with $n > 40$ or under Models A (for $\theta_{+2} - \theta_{2+}$) and C (for both parametric functions) with $n = 40$.

3.2 Findings

The effect of the priors on the posterior summaries of θ_{2+} practically disappears for $n \geq 400$ (Table 4). In addition, the means and CIs coincide, respectively, with the (point) HEIRs and the EUROs for such sample sizes (Table 5). These expected results, usually encountered in analyses of identifiable models, are not shared by the ones obtained for the non-identifiable parameters.

The posterior means, standard deviations (SD) and the CIs for $\theta_{+2} - \theta_{2+}$ and for log OR vary substantially across the different priors (Tables 6 and 8). Prior-dependence was also exhibited by Forster & Smith (1998) and Soares & Paulino (2001), but the fact that the level of dependence remains almost unaltered for $n \geq 4,000$ is alarming. Although the shifts of the posterior means are not very large when compared to the corresponding SDs, the sensitivity analysis of the priors suggests that one should not rely too much on this single summary. To deal with statistical ignorance, there will always be a set of non-observable (augmented) distributions compatible with the one we are able to observe. Posteriors will take us in the direction of one subset or another, depending on the chosen prior. Oversynthesizing in this scenario of lack of information may be dangerous. We believe that it is more appropriate to consider CIs.

Posterior SDs no longer tend to zero as the sample size increases for these non-identifiable parameters. On the contrary, they lie on plateaux that are drastically different across priors. These differences are more impressive for the priors commonly taken as non-informative. Of course, there is nothing wrong to use any of the priors suggested by the expert. Nevertheless, if we are looking for a really non-informative prior, there is no obvious choice, but we believe

Table 4: Posterior means, standard deviations (SD) and 95% equal-tailed credible intervals (CI) for θ_{2+}.

Prior 3	$\{c_{C_{ij}} = 0.1\}$			$\{c_{C_{ij}} = 0.25\}$			$\{c_{C_{ij}} = 0.5\}$			$\{c_{C_{ij}} = 1.0\}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	0.546	0.075	[0.398; 0.691]	0.546	0.075	[0.398; 0.691]	0.546	0.075	[0.399; 0.690]	0.546	0.075	[0.398; 0.691]
400	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.598]
4,000	0.550	0.008	[0.534; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]
40,000	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]
4,000,000	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]
Priors 3/2	$\{c_{C_{ij}} = 0.001\}$			$\{b_{ij} = a_{C_{ij}} = 0.1\}$			$\{b_{ij} = a_{C_{ij}} = 0.5\}$			$\{b_{ij} = a_{C_{ij}} = 1.0\}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	0.547	0.075	[0.398; 0.691]	0.550	0.077	[0.397; 0.698]	0.547	0.075	[0.398; 0.693]	0.546	0.074	[0.399; 0.689]
400	0.550	0.025	[0.501; 0.598]	0.550	0.025	[0.501; 0.599]	0.550	0.025	[0.501; 0.598]	0.549	0.025	[0.500; 0.597]
4,000	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.534; 0.565]
40,000	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]
4,000,000	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]
Prior 2	$\{b_{ij}\} = \text{Infl}$, $\{a_{C_{ij}} = 0.5\}$			$\{b_{ij}\} = \text{Inf10}$, $\{a_{C_{ij}} = 0.5\}$			$\{b_{ij} = 0.5\}$, $\{a_{C_{ij}}\} = \text{Infl}$			$\{b_{ij} = 0.5\}$, $\{a_{C_{ij}}\} = \text{Inf10}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	0.551	0.077	[0.399; 0.698]	0.560	0.070	[0.422; 0.694]	0.548	0.076	[0.397; 0.693]	0.547	0.076	[0.397; 0.693]
400	0.550	0.025	[0.502; 0.598]	0.551	0.025	[0.503; 0.599]	0.550	0.025	[0.501; 0.597]	0.550	0.025	[0.500; 0.598]
4,000	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.565]	0.550	0.008	[0.535; 0.566]
40,000	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]	0.550	0.002	[0.545; 0.555]
4,000,000	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]	0.550	0.000	[0.550; 0.550]

Table 5: HEIR and 95% strong and pointwise EUROs using maximum likelihood estimates for θ_{2+}.

	Models A, B and C - HEIR: [0.550; 0.550]
n	Strong and Pointwise
40	[0.396; 0.704]
400	[0.501; 0.599]
4,000	[0.535; 0.565]
40,000	[0.545; 0.555]
4,000,000	[0.550; 0.550]

Table 6: Posterior means, standard deviations (SD) and 95% equal-tailed credible intervals (CI) for $\theta_{+2} - \theta_{2+}$.

Prior 3	$\{c_{C_{ij}} = 0.1\}$			$\{c_{C_{ij}} = 0.25\}$			$\{c_{C_{ij}} = 0.5\}$			$\{c_{C_{ij}} = 1.0\}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	0.256	0.110	[0.029; 0.461]	0.256	0.107	[0.037; 0.455]	0.256	0.103	[0.046; 0.450]	0.256	0.099	[0.055; 0.445]
400	0.273	0.060	[0.159; 0.379]	0.273	0.055	[0.166; 0.373]	0.273	0.050	[0.174; 0.366]	0.273	0.045	[0.185; 0.357]
4,000	0.275	0.052	[0.190; 0.358]	0.275	0.047	[0.194; 0.354]	0.275	0.041	[0.202; 0.347]	0.275	0.034	[0.212; 0.337]
40,000	0.275	0.051	[0.197; 0.352]	0.275	0.046	[0.200; 0.350]	0.275	0.040	[0.205; 0.345]	0.275	0.032	[0.215; 0.335]
4,000,000	0.275	0.051	[0.200; 0.350]	0.275	0.046	[0.201; 0.349]	0.275	0.039	[0.205; 0.344]	0.275	0.032	[0.216; 0.334]
Priors 3/2	$\{c_{C_{ij}} = 0.001\}$			$\{b_{ij} = a_{C_{ij}} = 0.1\}$			$\{b_{ij} = a_{C_{ij}} = 0.5\}$			$\{b_{ij} = a_{C_{ij}} = 1.0\}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	0.256	0.113	[0.021; 0.465]	0.275	0.111	[0.045; 0.480]	0.270	0.105	[0.054; 0.467]	0.260	0.102	[0.053; 0.452]
400	0.273	0.064	[0.153; 0.385]	0.277	0.060	[0.162; 0.382]	0.281	0.051	[0.179; 0.372]	0.282	0.045	[0.190; 0.366]
4,000	0.275	0.057	[0.186; 0.361]	0.277	0.052	[0.190; 0.359]	0.282	0.041	[0.204; 0.350]	0.284	0.034	[0.217; 0.342]
40,000	0.275	0.056	[0.196; 0.354]	0.277	0.051	[0.198; 0.352]	0.282	0.040	[0.207; 0.346]	0.284	0.032	[0.220; 0.339]
4,000,000	0.275	0.056	[0.200; 0.350]	0.277	0.051	[0.200; 0.350]	0.282	0.040	[0.207; 0.346]	0.284	0.032	[0.220; 0.339]
Prior 2	$\{b_{ij}\} = \text{Inf1}$, $\{a_{C_{ij}} = 0.5\}$			$\{b_{ij}\} = \text{Inf10}$, $\{a_{C_{ij}} = 0.5\}$			$\{b_{ij} = 0.5\}$, $\{a_{C_{ij}}\} = \text{Inf1}$			$\{b_{ij} = 0.5\}$, $\{a_{C_{ij}}\} = \text{Inf10}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	0.291	0.104	[0.075; 0.486]	0.271	0.093	[0.082; 0.448]	0.253	0.108	[0.031; 0.455]	0.289	0.099	[0.089; 0.475]
400	0.289	0.050	[0.185; 0.377]	0.286	0.049	[0.185; 0.372]	0.265	0.055	[0.160; 0.367]	0.299	0.037	[0.222; 0.368]
4,000	0.289	0.040	[0.208; 0.352]	0.287	0.040	[0.207; 0.350]	0.265	0.046	[0.191; 0.349]	0.299	0.022	[0.248; 0.337]
40,000	0.289	0.039	[0.210; 0.348]	0.287	0.039	[0.209; 0.347]	0.265	0.046	[0.198; 0.347]	0.299	0.020	[0.253; 0.331]
4,000,000	0.289	0.039	[0.210; 0.347]	0.287	0.039	[0.209; 0.347]	0.265	0.045	[0.200; 0.347]	0.299	0.020	[0.254; 0.330]

Table 7: HEIRs and 95% strong, pointwise and weak EUROs using maximum likelihood estimates for $\theta_{+2} - \theta_{2+}$.

	Model B - HEIR: [0.244; 0.344]			Models A and C - HEIR: [0.200; 0.350]		
n	Strong	Pointwise	Weak	Strong	Pointwise	Weak
40	[0.027; 0.522]	[0.056; 0.498]	[0.073; 0.484]	[−0.010; 0.527]	[0.022; 0.500]	[0.052; 0.475]
400	[0.175; 0.400]	[0.186; 0.391]	[0.208; 0.373]	[0.134; 0.406]	[0.144; 0.397]	[0.173; 0.373]
4,000	[0.222; 0.362]	[0.226; 0.359]	[0.240; 0.347]	[0.179; 0.368]	[0.182; 0.365]	[0.200; 0.350]
40,000	[0.237; 0.349]	[0.238; 0.348]	[0.246; 0.342]	[0.193; 0.356]	[0.194; 0.355]	[0.204; 0.347]
4,000,000	[0.243; 0.344]	[0.243; 0.344]	[0.246; 0.341]	[0.199; 0.351]	[0.199; 0.350]	[0.204; 0.347]

Table 8: Posterior means, standard deviations (SD) and 95% equal-tailed credible intervals (CI) for log OR.

Prior 3	$\{c_{C_{ij}} = 0.1\}$			$\{c_{C_{ij}} = 0.25\}$			$\{c_{C_{ij}} = 0.5\}$			$\{c_{C_{ij}} = 1.0\}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	−0.02	1.15	[−2.25; 2.29]	−0.04	1.07	[−2.10; 2.11]	−0.06	0.97	[−1.95; 1.90]	−0.06	0.88	[−1.78; 1.68]
400	−0.04	0.77	[−1.39; 1.36]	−0.05	0.69	[−1.30; 1.25]	−0.07	0.61	[−1.18; 1.12]	−0.08	0.51	[−1.02; 0.93]
4,000	−0.04	0.72	[−1.18; 1.13]	−0.05	0.64	[−1.12; 1.07]	−0.07	0.55	[−1.03; 0.97]	−0.08	0.45	[−0.89; 0.80]
40,000	−0.04	0.71	[−1.12; 1.07]	−0.05	0.63	[−1.09; 1.04]	−0.07	0.55	[−1.02; 0.96]	−0.08	0.44	[−0.87; 0.79]
4,000,000	−0.04	0.71	[−1.10; 1.05]	−0.05	0.63	[−1.09; 1.04]	−0.07	0.54	[−1.02; 0.96]	−0.08	0.44	[−0.87; 0.79]
Priors 3/2	$\{c_{C_{ij}} = 0.001\}$			$\{b_{ij} = a_{C_{ij}} = 0.1\}$			$\{b_{ij} = a_{C_{ij}} = 0.5\}$			$\{b_{ij} = a_{C_{ij}} = 1.0\}$		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	−0.01	1.24	[−2.40; 2.46]	−0.04	1.27	[−2.50; 2.53]	−0.01	1.10	[−2.15; 2.21]	0.00	0.98	[−1.92; 1.95]
400	−0.03	0.84	[−1.47; 1.44]	−0.02	0.78	[−1.41; 1.37]	−0.02	0.63	[−1.19; 1.16]	−0.02	0.54	[−1.04; 1.02]
4,000	−0.03	0.79	[−1.21; 1.17]	−0.02	0.72	[−1.18; 1.13]	−0.02	0.56	[−1.03; 0.98]	−0.02	0.46	[−0.88; 0.83]
40,000	−0.03	0.78	[−1.13; 1.09]	−0.02	0.71	[−1.12; 1.07]	−0.02	0.55	[−1.02; 0.97]	−0.03	0.45	[−0.87; 0.82]
4,000,000	−0.03	0.78	[−1.10; 1.05]	−0.03	0.71	[−1.10; 1.05]	−0.02	0.55	[−1.02; 0.97]	−0.03	0.45	[−0.87; 0.82]
Prior 2	$\{b_{ij}\}$ = Infl, $\{a_{C_{ij}} = 0.5\}$			$\{b_{ij}\}$ = Infl0, $\{a_{C_{ij}} = 0.5\}$			$\{b_{ij} = 0.5\}$, $\{a_{C_{ij}}\}$ = Infl			$\{b_{ij} = 0.5\}$, $\{a_{C_{ij}}\}$ = Infl0		
n	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI	Mean	SD	95% CI
40	0.03	1.22	[−2.36; 2.47]	0.14	0.98	[−1.77; 2.08]	0.01	1.10	[−2.11; 2.24]	0.14	1.06	[−1.96; 2.28]
400	0.02	0.63	[−1.17; 1.19]	0.06	0.61	[−1.13; 1.18]	0.01	0.68	[−1.24; 1.30]	0.19	0.43	[−0.68; 1.02]
4,000	0.02	0.56	[−1.03; 1.00]	0.05	0.55	[−1.00; 0.99]	0.01	0.63	[−1.08; 1.08]	0.21	0.31	[−0.46; 0.76]
40,000	0.02	0.55	[−1.01; 0.97]	0.05	0.54	[−0.99; 0.98]	0.01	0.62	[−1.07; 1.05]	0.21	0.30	[−0.43; 0.72]
4,000,000	0.02	0.55	[−1.01; 0.97]	0.05	0.54	[−0.99; 0.98]	0.01	0.62	[−1.07; 1.04]	0.21	0.29	[−0.42; 0.72]

Table 9: HEIRs and 95% strong, pointwise and weak EUROs using maximum likelihood estimates for log OR.

	Model A - HEIR: [−0.27; 0.22]			Model B - HEIR: [−0.97; 0.36]			Model C - HEIR: [−1.10; 1.05]		
n	Strong	Pointwise	Weak	Strong	Pointwise	Weak	Strong	Pointwise	Weak
40	[−1.73; 2.29]	[−1.59; 2.10]	[−1.54; 2.02]	[−2.72; 2.43]	[−2.45; 2.12]	[−2.24; 1.87]	[−2.84; 2.88]	[−2.56; 2.59]	[−2.14; 2.15]
400	[−0.73; 0.88]	[−0.66; 0.77]	[−0.58; 0.66]	[−1.52; 1.01]	[−1.43; 0.91]	[−1.21; 0.65]	[−1.65; 1.63]	[−1.56; 1.54]	[−1.25; 1.21]
4,000	[−0.42; 0.43]	[−0.39; 0.40]	[−0.33; 0.30]	[−1.14; 0.56]	[−1.11; 0.53]	[−0.98; 0.37]	[−1.27; 1.23]	[−1.24; 1.20]	[−1.07; 1.02]
40,000	[−0.32; 0.29]	[−0.31; 0.28]	[−0.27; 0.22]	[−1.02; 0.42]	[−1.01; 0.41]	[−0.94; 0.32]	[−1.15; 1.11]	[−1.14; 1.10]	[−1.05; 1.00]
4,000,000	[−0.28; 0.23]	[−0.28; 0.23]	[−0.26; 0.21]	[−0.97; 0.36]	[−0.97; 0.36]	[−0.93; 0.32]	[−1.10; 1.06]	[−1.10; 1.05]	[−1.05; 0.99]

that an adequate one is certainly far from those with hyper-parameters equal to 0.5 or 1.0 usually chosen for the Dirichlet family. We repeated the analysis based on Prior 3 with increasingly diffuse priors until stable results were obtained for $n = 4,000,000$, and they did not change further for hyper-parameters with values smaller than 0.001. In the same spirit, one should be aware that even priors conventionally considered to be slightly informative may actually change the posterior distributions considerably for any sample size.

Priors with all hyper-parameters smaller than 1.0 exhibited multi-modal posterior densities for $\theta_{+2} - \theta_{2+}$ and for $\log \mathrm{OR}$ when $n \geq 4,000$; see Figure 1, where some posterior densities obtained from Prior 3 are displayed. The same was observed for $\boldsymbol{\theta}$, when $n \geq 400$, and for $\boldsymbol{\lambda}$, irrespectively of the sample sizes (results not shown). In the continuous data case, Scharfstein et al. (2003) pointed out that this may also happen when informative priors for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ are not simultaneously 'compatible' with the observed data, and suggested that "the experts re-evaluate their priors by reducing the 'informativeness' of one of them" to avoid the unsatisfactory multi-modality for inferences. We believe that uni-modality should not be expected for posterior densities of non-identifiable parameters. Actually, multi-modality may be the rule rather than the exception in our case, given that the posterior distribution of $\boldsymbol{\theta}$ can be expressed as a finite mixture of Dirichlet distributions. Therefore, this is an inherent characteristic that should be taken into consideration and not simply avoided.

Results from classical inferences using the parsimonious Model A coincide with those based on the unconstrained Model C for $\theta_{+2} - \theta_{2+}$, but not for $\log \mathrm{OR}$ (Tables 7 and 9). This shows that for different parametric functions, over-parameterized missingness models may affect inferences in various ways. Taking, for example, the OR into account, it is clear that the presence of the interaction between Y_1 and Y_2 in the missingness model is more important than the presence of the main effect of Y_2. This leads to the ad hoc non-hierarchical Model B. The HEIR for $\log \mathrm{OR}$ based on Model B is considerably larger than the one resulting from Model A, but its upper limit is still too far from the one obtained for Model C. On the other hand, the lower limit of the HEIR for $\theta_{+2} - \theta_{2+}$ based on Model

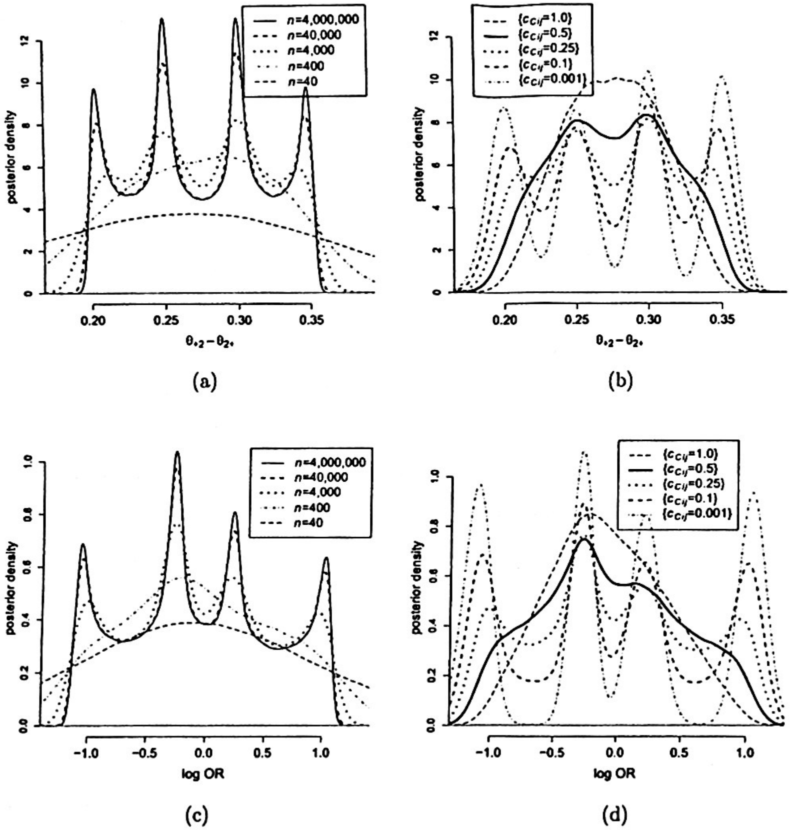

Figure 1: Posterior densities of $\theta_{+2} - \theta_{2+}$, (a) and (b), and log OR, (c) and (d), using Prior 3: $\{c_{Cij} = 0.25\}$ in (a) and (c), and $n = 4,000$ in (b) and (d).

B is also distant from the corresponding ones obtained from Models A and C. The shorter intervals obtained under Models A and B are welcome, but only if these models make sense to the experts. Otherwise, they may generate misleading conclusions.

Comparisons between Bayesian and classical approaches make sense only when restricted to the less informative priors in the former and to the unconstrained Model C in the latter. The CIs are, in general, somewhat closer to the pointwise EUROs when the values of the hyper-parameters approach zero. The difference between both intervals decreases as the sample size increases.

It is natural to search for a more parsimonious missingness model, since this leads to more precise inferences. The so-called missing at random (MAR) mechanism is usually the focus of attention not only because in many problems it makes sense to assume that the missing data depend only on the observed data, but also by virtue of the ignorability of the corresponding parameter $\boldsymbol{\lambda}$ from the standpoint of likelihood-based or Bayesian inferences on $\boldsymbol{\theta}$ when both kinds of parameters are not functionally or a priori related (Rubin, 1976). In our case, the MAR model may be obtained by taking $\lambda_{\{11,12\}(11)} = \lambda_{\{11,12\}(12)}$ and $\lambda_{\{21,22\}(21)} = \lambda_{\{21,22\}(22)}$. Comparing the posterior distributions of these individual parameters as in Soares & Paulino (2001) may misleadingly point to a rejection of the MAR hypothesis, while their differences indicate a fairer picture. This is clear in Figure 2, where we display boxplots of the posterior draws of the parameters, and of differences thereof, for the more informative Prior 2 hyper-parameters with $n = 4,000,000$. The MAR model was not rejected a posteriori for any of the prior distributions and sample sizes. The MAR assessment can also be conducted via the classical approach if $\boldsymbol{\lambda}$ is switched to the estimable parameter set. However, it is not possible to distinguish between MAR and non-MAR models with information obtained exclusively from the data, as shown, for example, by Molenberghs et al. (2008). This notwithstanding, distinctions may occur if more informative priors or shorter ranges for $\boldsymbol{\Omega}$ are used, and/or constrained models are adopted.

For the same reasons why the MAR hypothesis cannot be accepted or rejected on the basis of the present results, we cannot assert whether there is an association between Y_1 and Y_2 or not. In practice, non-rejections are usually taken as evidence in favour of the null hypotheses. The use of hypothesis tests without looking at the power for different alternatives may not be a so condemnable procedure for analyses of regular identifiable models when the power

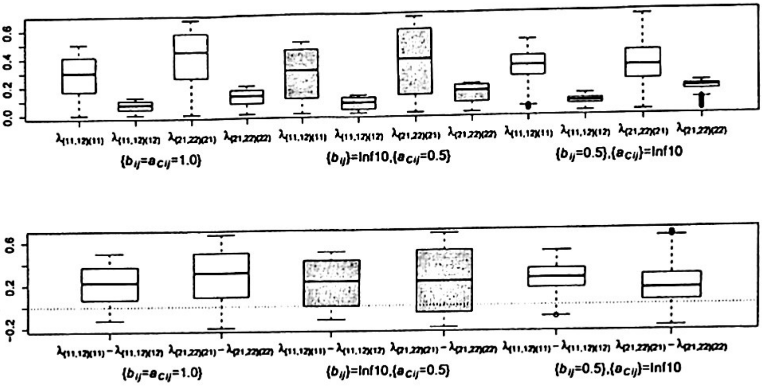

Figure 2: Boxplots of some λ parameters a posteriori, and differences thereof, for Prior 2 with $n = 4,000,000$.

of the test is expected to be high. Nevertheless, this is more dangerous when the hypothesis involves non-identifiable parameters. By definition, tests obtained via EUROs have very low power for most of the alternatives that belong to the ignorance region, irrespectively of the sample size. Similarly, CIs based on priors for which the hyper-parameters have values close to zero are also generally useless to reject a hypothesis if the alternative lies within the non-parametric ignorance interval. These wide intervals have an appealing protection against unverifiable assumptions, but they deserve careful evaluation.

Although we address a series of important issues for the analyses of non-identifiable models, rejection of the marginal homogeneity (or symmetry) and other inferential conclusions remain unchanged for all priors and models. In the next section, we present an alternative motivating scenario.

4 Reanalysis of the Data from the Collaborative Perinatal Project

We reanalyse the data of Table 1 along the lines of the preceding section. The results obtained via Bayesian and classical approaches are displayed, respectively, in Tables 10 and 11. The classical sensitivity analysis with an unconstrained missingness model is conducted following the guidelines of Section 2.4.

Table 10: Posterior means, standard deviations (SD) and 95% equal-tailed credible intervals (CI) for OR.

Prior	Hyper-parameters	Mean	SD	95% CI
3	$\{c_{Cij} = 0.001\}$	1.49	0.40	[0.85; 2.40]
3	$\{c_{Cij} = 0.1\}$	1.48	0.35	[0.90; 2.27]
3	$\{c_{Cij}\} = ^{*}$	1.47	0.33	[0.93; 2.19]
3	$\{c_{Cij} = 0.5\}$	1.45	0.26	[1.02; 2.01]
3	$\{c_{Cij} = 1.0\}$	1.44	0.21	[1.09; 1.89]
2	$\{b_{ij} = a_{Cij} = 0.1\}$	1.48	0.35	[0.90; 2.28]
2	$\{b_{ij} = a_{Cij} = 0.5\}$	1.46	0.26	[1.03; 2.02]
2	$\{b_{ij} = a_{Cij} = 1.0\}$	1.46	0.21	[1.10; 1.90]

* $c_{Cij} = 0.125$, if $C = \{11, 12, 21, 22\}$, and $c_{Cij} = 0.25$, otherwise.

Table 11: HEIR and 95% strong, pointwise and weak EUROs using maximum likelihood estimates for OR.

Model	Ω	HEIR	Strong	Pointwise	Weak
1	$[-3;\ 3]$	[1.06; 2.04]	[1.01; 2.14]	[1.02; 2.12]	[1.07; 2.02]
1	$[-5;\ 5]$	[0.94; 2.23]	[0.90; 2.33]	[0.91; 2.32]	[0.96; 2.19]
1	$[-10; 10]$	[0.89; 2.32]	[0.84; 2.45]	[0.85; 2.43]	[0.90; 2.28]
2	$[-10; 10]^2$	[1.10; 1.73]	[1.06; 1.81]	[1.07; 1.80]	[1.11; 1.73]
Unconstrained	$[-10; 10]^7$	[0.82; 2.50]	[0.79; 2.61]	[0.79; 2.59]	[0.84; 2.44]

In addition to our previous discussion, we note that even the CI based on the least informative prior (Prior 3 with $\{c_{Cij} = 0.001\}$) is now closer to the weak EURO than the pointwise EURO for the results obtained with the unconstrained missingness model. The problem here is the possibility of obtaining different conclusions depending on the subjective choices within each approach. Unless some prior information suggests that it is reasonable to restrict inferences to Model 2 or to Model 1 with $\Omega = [-3; 3]$, or to use an informative prior distribution (or "non-informative" ones for which the hyper-parameters have values ≥ 0.5), we cannot conclude whether or not maternal smoking and low newborn weight are associated. These conclusions are different from those of Baker et al. (1992), who conducted an informal sensitivity analysis based on 9 models that reinforced their belief in a significant association. Kenward et al. (2001) presented another example wherein an informal sensitivity analysis that led to consistent conclusions was misleading, but it is interesting to note that this holds when the sample size is as large as $n = 57,061$ and the proportion of data that are missing is as small as 7%.

5 Concluding Remarks

In an effort to include relevant aspects of a problem into its statistical analysis, one may frequently arrive at non-identifiable models. We explored Bayesian and classical approaches to analyse these specialized models in a missing categorical response setting. Both frameworks provide sensible answers if meticulous attention is devoted to the subjective choices and there is a clear understanding of the strengths and limitations of the analyses. Otherwise, one may generate misleading conclusions.

The keystone to understand analyses of over-parameterized models is related to the distinction between observable and non-observable data, or what we can and cannot learn from the study. At first sight, simplifying a model to gain identifiability seems to be a good decision that throws us directly to what can be learned from the study and prevents us from dealing with the obscure non-observable part. Indeed, to a certain degree, inferences obtained therefrom are much less uncertain, but rely on usually untestable underlying assumptions. Thus, identifiable models turn out to be an extreme and subjective tool to deal with the ignorance about the non-observable data.

In the process of facing statistical ignorance and performing analyses of non-identifiable models, new issues have surfaced. Particularly, we learned that larger sample sizes do not increase our knowledge about what is non-identifiable when statistical imprecision is already sufficiently negligible. Asymptotically, credible and uncertainty intervals cover the ignorance interval to a great extent. Therefore, the former intervals are useless if we want to distinguish between the parameter values that belong to the latter interval. As a consequence, if the alternative lies in the ignorance interval, the power of these tests no longer tends to one as the sample size increases. This is the reason why it is inadmissible to consider the non-rejection of a hypothesis as its acceptance, in general. This should always be clarified in any report of these analyses.

An objective strategy is to use unconstrained missingness models, diffuse prior distributions in Bayesian analyses or try to cover the whole parameter space of the sensitivity parameter in the classical approach. Although this constitutes a reasonable starting point that, in many cases, seems to be the only alternative, we believe that a more suitable analysis should attempt to incorporate all the available information. This may be achieved by following the opposite route, i.e., considering informative prior distributions, or plausible ranges for the sensitivity parameter, and/or constrained missingness models. Our objective here is only to illustrate that such added information should be carefully evaluated, because arbitrary choices may have a significant impact on results.

Under a Bayesian perspective, we showed that priors usually taken as slightly informative or non-informative (e.g., by setting the hyper-parameters of Dirichlet distributions equal to 0.5 or 1.0) may actually add substantial information, irrespectively of the sample size. Constrained over-parameterized models may also affect the results considerably. Although such models may also be used in Bayesian analyses, they are more frequently adopted in the classical approach, mainly because it is cumbersome to deal with too many sensitivity parameters. In the reanalysis of the Collaborative Perinatal Project, for example, we considered only 3 values, $\{-10, 0, 10\}$, for each of the 7 sensitivity parameters, obtaining

$3^7 = 2187$ combinations. Using such a coarse grid could have generated an even shorter ignorance region than the ones obtained with constrained missingness models, but it was not difficult, for this particular contingency table, to identify the configurations that lead to extreme inferences for the odds ratio and, as a consequence, we could validate the resulting ignorance region of the unconstrained model.

As the credible intervals obtained with diffuse priors were somewhat close to the uncertainty intervals obtained with the unconstrained missingness models, the Bayesian approach may offer advantages over the classical sensitivity analysis when (a) there is prior information to be incorporated, or (b) no prior information is available, but there is interest in results for a high-dimensional unconstrained missingness model. In such a case, however, we must point that diffuse (proper) priors are convenient to avoid averaging out the extremes of the credible intervals, but these intervals may still be a little narrower than the weak uncertainty intervals that would be obtained otherwise.

While Prior 2 is the more appropriate one to incorporate prior information, the generated Gibbs sampling chains become highly autocorrelated as the sample size increases and the values of hyper-parameters approach zero. Computational time for some of the analyses ranged from hours to weeks, depending on whether $n \leq 4000$, $n = 40,000$ or $4,000,000$. Nevertheless, we may use an ordinary Monte Carlo approach with Prior 3 and conduct analyses for any hyper-parameter values in a few minutes. Therefore, when there is no prior information, we suggest to use Jeffreys' marginal prior for the identifiable parameters with the values of hyper-parameters associated to the prior distribution of the non-identifiable parameters close to zero. Note that in continuous data settings, looking for non-informative priors is even trickier (Daniels & Hogan, 2007).

When attempting to incorporate available information about the missing data generating mechanism, not only covariates may be useful, but also some justification for the missing values may facilitate the task of including certain assumptions, as MAR or non-MAR, for some sampling units. In longitudinal studies, time is generally an additional important factor. When the missingness pattern is monotone, it may be reasonable to ensure that the dropouts do not depend upon future values (Diggle & Kenward, 1994; Kenward et al., 2003), leading to an interesting and more parsimonious over-parameterized model than the unconstrained one.

Acknowledgements

We gratefully acknowledge the financial supports to this research: Frederico Z. Poleto and Julio M. Singer, from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil; Carlos Daniel Paulino, from Fundação para a Ciência e Tecnologia (FCT) through the research centre CEAUL-FCUL, Portugal; Geert Molenberghs, from the IAP research Network P6/03 of the Belgian Government (Belgian Science Policy).

References

Agresti, A. & Min, Y. (2005). Frequentist performance of Bayesian confidence intervals for comparing proportions in 2 × 2 contingency tables. *Biometrics*, **61**, 515–523.

Baker, S. G., Rosenberger, W. F. & DerSimonian, R. (1992). Closed-form estimates for missing counts in two-way contingency tables. *Statist. Med.*, **11**, 643–657.

Bedrick, E. J., Christensen, R. & Johnson, W. (1996). A new perspective on priors for generalized linear models. *J. Amer. Statist. Assoc.*, **91**, 1450–1460.

Copas, J. B. & Li, H. G. (1997). Inference for non-random samples (with discussion). *J. R. Stat. Soc. B*, **59**, 55–95.

Daniels, M. J. & Hogan, J. W. (2007). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis.* London: Chapman & Hall.

Dickey, J. M. (1983). Multiple hypergeometric functions: probabilistic interpretations and statistical uses. *J. Amer. Statist. Assoc.*, **78**, 628–637.

Diggle, P. J. & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Appl. Statist.*, **43**, 49–93.

Fay, R. E. (1986). Causal models for patterns of nonresponse. *J. Amer. Statist. Assoc.*, **81**, 354–365.

Forster, J. J. & Smith, P. W. F. (1998). Model-based inference for categorical survey data subject to non-ignorable non-response (with discussion). *J. R. Stat. Soc. B*, **60**, 57–70, 89–102.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In *Bayesian Statistics 4*. Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith. Oxford: Oxford University Press, 169–193.

Glynn, R. J., Laird, N. M. & Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse (with discussion). In *Drawing Inferences from Self-Selected Samples*. Ed. H. Wainer. Mahwah: Lawrence Erlbaum Associates, 115–151.

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables (with discussion). *Stat. Sci.*, **20**, 111–140.

Heidelberger, P. & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Oper. Res.*, **31**, 1109–1144.

Jiang, T. J. & Dickey, J. M. (2008). Bayesian methods for categorical data under informative censoring. *Bayesian Anal.*, **3**, 541–554.

Kenward, M. G., Goetghebeur, E. & Molenberghs, G. (2001). Sensitivity analysis for incomplete categorical data. *Stat. Model.*, **1**, 31–48.

Kenward, M. G., Molenberghs, G. & Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika*, **90**, 53–71.

Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. New York: John Wiley & Sons.

Molenberghs, G., Beunckens, C., Sotto, C. & Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *J. R. Stat. Soc. B*, **70**, 371–388.

Molenberghs, G., Goetghebeur, E., Lipsitz, S. R. & Kenward, M. G. (1999). Nonrandom missingness in categorical data: strengths and limitations. *Amer. Statist.*, **53**, 110–118.

Molenberghs, G., Kenward, M. G. & Goetghebeur, E. (2001). Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case. *Appl. Statist.*, **50**, 15–29.

Neal, R. M. (2003). Slice sampling (with discussion). *Ann. Statist.*, **31**, 705–767.

Neath, A. A. & Samaniego, F. J. (1997). On the efficacy of Bayesian inference for nonidentifiable models. *Amer. Statist.*, **51**, 225–232.

Nordheim, E. V. (1984). Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *J. Amer. Statist. Assoc.*, **79**, 772–780.

Paulino, C. D. & Pereira, C. A. B. (1992). Bayesian analysis of categorical data informatively censored. *Comm. Statist. Theory Methods*, **21**, 2689–2705.

Paulino, C. D. & Pereira, C. A. B. (1994). On identifiability of parametric statistical models. *Stat. Methods Appl.*, **3**, 125–151.

Paulino, C. D. & Pereira, C. A. B. (1995). Bayesian methods for categorical data under informative general censoring. *Biometrika*, **82**, 439–446.

Raftery, A. E. & Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*. Eds. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith. Oxford: Oxford University Press, 763–773.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

Scharfstein, D. O., Daniels, M. J. & Robins, J. M. (2003). Incorporating prior beliefs about selection bias into the analysis of randomized trials with missing outcomes. *Biostatistics*, **4**, 495–512.

Scharfstein, D. O., Rotnitzky, A. & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *Journal of the American Statistical Association*, **94**, 1096–1146.

Soares, P. (2004). *Bayesian analysis of coarse categorical data (in Portuguese)*. Ph.D. thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.

Soares, P. & Paulino, C. D. (2001). Incomplete categorical data analysis: a Bayesian perspective. *J. Stat. Comput. Simul.*, **69**, 157–170.

Soares, P. & Paulino, C. D. (2007). Log-linear models for coarse categorical data. In *Proceedings of the 56th Session of the International Statistical Institute*. Eds. M. I. Gomes, D. Pestana & P. Silva.

Tanner, M. A. & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.*, **82**, 528–550.

Tian, G.-L., Ng, K. W. & Geng, Z. (2003). Bayesian computation for contingency tables with incomplete cell-counts. *Statist. Sinica*, **13**, 189–206.

Vansteelandt, S., Goetghebeur, E., Kenward, M. G. & Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statist. Sinica*, **16**, 953–979.

Walker, S. (1996). A Bayesian maximum a posteriori algorithm for categorical data under informative general censoring. *The Statistician*, **45**, 293–298.

Wang, C.-P. & Ghosh, M. (2000). Bayesian analysis of bivariate competing risks models. *Sankhya*, **62**, 388–401.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2010-01 - BUENO, V.C. Component importance for a coherent system under a hyperminimal distribution. 09p. (RT-2010-01)

2010-02 - ABADI, M., ESTEVES, L.G., SIMONIS, A. Pointwise approximationos for sums of non-identically distributed Bernoulli trials. 12p. (RT-MAE-2010-02)

2010-03 - BUENO, V.C. A Series Representation of a Coherent System. 16p. (RT-MAE-2010-03)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05314-970 - São Paulo, Brasil