

Sensitivity analysis for incomplete continuous data

Peer-reviewed author version

Poleto, Frederico Z.; MOLENBERGHS, Geert; Paulino, Carlos Daniel & Singer, Julio M. (2011) Sensitivity analysis for incomplete continuous data. In: TEST, 20(3), p. 589-606.

DOI: 10.1007/s11749-010-0219-x

Handle: <http://hdl.handle.net/1942/12928>

Sensitivity analysis for incomplete continuous data

Frederico Z. Poletto¹, Geert Molenberghs²,
Carlos Daniel Paulino³ and Julio M. Singer¹

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

²I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium,
and Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

³Instituto Superior Técnico, Universidade Técnica de Lisboa (and CEAUL-FCUL),
Av. Rovisco Pais, Lisboa, 1049-001, Portugal

Abstract

In studies with missing data, statisticians typically identify the model via necessarily untestable assumptions and then perform sensitivity analyses to assess their effect on the conclusions. Both the parameterization and the identification of the model play an important role in translating the assumptions to non-statisticians and, consequently, in obtaining relevant information from experts or historical data. Specifically for continuous data, much of the earlier work has been developed under the assumption of normality and/or with hard-to-interpret sensitivity parameters. We derive a simple approach for estimating means, standard deviations and correlations that avoids parametric distributional assumptions for the outcomes. Adopting a pattern-mixture model parameterization, we use non-identifiable means, standard deviations, correlations or functions thereof as sensitivity parameters, which are more easily elicited.

Key words: Identifiability; Ignorance interval; Missing data; Pattern-mixture model; Uncertainty interval.

1 Introduction

In many problems with missing data, untestable assumptions, such as missing at random (MAR), following the terminology of Rubin (1976), are required to identify appropriate statistical models. Such assumptions are usually questionable and statisticians commonly bypass the problem via sensitivity analyses. Specifically for continuous data, Rubin (1977), Little (1994), Little & Wang (1996) and Daniels & Hogan (2000) propose sensitivity analyses under assumptions of normality, while Rotnitzky, Robins & Scharfstein (1998), Scharfstein, Rotnitzky & Robins (1999) and Rotnitzky et al. (2001) use inverse probability weighted (IPW) methods in the context of semi-parametric models for similar purposes. Reviews of some of these and other approaches are presented in Fitzmaurice et al. (2008, Chs. 18, 20, and 22).

Although such methodological developments are useful in many situations, there are cases where they may be difficult to apply. To bypass the problem in cases where the interest lies in estimating means, standard deviations and correlations under minor and meaningful assumptions, we combine some of these ideas and derive a simple approach. In this context, we adopt a pattern-mixture model parameterization (Glynn, Laird & Rubin, 1986; Little & Rubin, 2002) and employ non-identifiable means, standard deviations, and correlations, or functions thereof, as sensitivity parameters. This strategy is similar to the one adopted by Daniels & Hogan (2000), although we do not assume any parametric distribution for the outcomes. We believe that, in many applications, it may be easier to elicit information on these sensitivity parameters than on the selection-bias functions used by Rotnitzky and colleagues. Instead of IPW methods, we simply estimate the identifiable parameters by their sample analogues.

In Section 2, we present the data on American colleges that will be used to illustrate the methods described in the remainder of the paper. We introduce the ideas in a univariate setup in Section 3 and consider a multivariate extension in Section 4.

2 The American colleges data

The U.S. News & World Report's Guide to America's Best Colleges 1995 collected more than 30 variables encompassing characteristics such as admission, costs, infrastructure, and performance of the students on 1,302 American colleges and universities. Allison (2001) considered the estimation of means and standard deviations for 7 such variables under the MAR assumption using the EM algorithm. For the sake of our exposition, it suffices to focus only on 3 of them, namely, CSAT (average combined math and verbal Scholastic Assessment Test), GRADRAT (ratio between the number of graduating seniors and the number of enrolled students four years earlier $\times 100$) and an indicator of public versus private colleges. One college had a GRADRAT equal to 118 and, therefore, the corresponding value was considered missing. The public-private college administration indicator was the only variable without missing values. We are interested in two questions: i) do the public and private colleges have different mean CSAT? and ii) are CSAT and GRADRAT linearly correlated? Descriptive statistics are displayed in Tables 1 and 2. Because all American colleges matching the criteria adopted in the study were surveyed, the data make up the entire study population and therefore, standard errors and confidence intervals will be computed and discussed merely for illustrative and instructive purposes.

Table 1: Counts, means and standard deviations (SD) for CSAT

College administration	CSAT observed			CSAT missing		
	Count	Mean	SD	Count	Mean	SD
Public	251	945.3	107.5	219	?	?
Private	528	978.8	129.2	304	?	?

? denotes non-observed values

3 Univariate case

Let Y_i denote the measurement on the i -th unit of the study and R_i be an indicator variable taking on the value 1 if Y_i is observed and 0, otherwise, $i = 1, \dots, n$. Assume,

Table 2: Counts, means, standard deviations (SD) and correlations for GRADRAT and CSAT

Missingness pattern		Count	GRADRAT		CSAT		Correlation between GRADRAT and CSAT
GRADRAT	CSAT		Mean	SD	Mean	SD	
Observed	Observed	731	62.0	18.5	974.0	123.0	0.594
Observed	Missing	472	57.7	19.0	?	?	?
Missing	Observed	48	?	?	876.6	93.5	?
Missing	Missing	51	?	?	?	?	?

? denotes non-observed values

further that $n_1 = \sum_{i=1}^n R_i$ units are observed and $n_0 = n - n_1$ units are missing. Using the pattern-mixture model parameterization and conditional expectation properties, we have

$$\mu = E(Y_i) = \gamma_1 \mu_{(1)} + \gamma_0 \mu_{(0)}, \quad (1)$$

$$\sigma^2 = \text{Var}(Y_i) = \gamma_1 \sigma_{(1)}^2 + \gamma_0 \sigma_{(0)}^2 + \gamma_1 (\mu_{(1)} - \mu)^2 + \gamma_0 (\mu_{(0)} - \mu)^2, \quad (2)$$

where $\gamma_r = P(R_i = r)$, $\mu_{(r)} = E(Y_i | R_i = r)$ and $\sigma_{(r)}^2 = \text{Var}(Y_i | R_i = r)$, for $r = 0, 1$.

We can estimate γ_1 ($\gamma_0 = 1 - \gamma_1$), $\mu_{(1)}$ and $\sigma_{(1)}^2$ by their sample counterparts $\hat{\gamma}_1$ ($\hat{\gamma}_0$), $\hat{\mu}_{(1)}$ and $\hat{\sigma}_{(1)}^2$. The parameters $\mu_{(0)}$ and $\sigma_{(0)}^2$ are not identified from the observed data, but, if we set values for them, we may obtain an unbiased estimate $\hat{\mu}(\mu_{(0)})$ of $\mu(\mu_{(0)})$ and a consistent estimate $\hat{\sigma}^2(\mu_{(0)}, \sigma_{(0)}^2)$ of $\sigma^2(\mu_{(0)}, \sigma_{(0)}^2)$. In this setup, $\boldsymbol{\omega} = \mu_{(0)}$ or $\boldsymbol{\omega} = (\mu_{(0)}, \sigma_{(0)}^2)$ are the so-called sensitivity parameters for the purpose of estimating μ or σ^2 , respectively.

The range of estimates obtained after repeating the analysis over a set $\boldsymbol{\Omega}$ of values for $\boldsymbol{\omega}$ provides a Honestly Estimated Ignorance Region (HEIR). Likewise, the union of $100(1 - \alpha)\%$ confidence regions obtained for different values of $\boldsymbol{\omega}$ provides a $100(1 - \alpha)\%$ Estimated Uncertainty Region (EURO). In the same way that standard errors and confidence regions quantify statistical imprecision stemming from sampling, ignorance regions measure the statistical ignorance on account of deficiencies of the observation process, like missing data; likewise, the uncertainty region assesses the statistical uncertainty caused by the combination of imprecision and ignorance. Vansteelandt et al. (2006) consider a formal approach

to the problem and provide appropriate definitions of consistency and coverage for these regions. They show how to construct EUROs with uncertainty level $100(1 - \alpha)\%$ for a scalar parameter π according to each definition of the uncertainty region: (i) strong EUROs cover $\pi(\boldsymbol{\omega})$ simultaneously for all $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ with at least $100(1 - \alpha)\%$ probability, (ii) pointwise EUROs cover $\pi(\boldsymbol{\omega})$ uniformly over $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ with at least $100(1 - \alpha)\%$ probability, and (iii) weak EUROs have an expected overlap with the ignorance region of at least $100(1 - \alpha)\%$. Strong EUROs are conservative pointwise EUROs, which in turn, are conservative weak EUROs. The choice among the three versions of EUROs depends on which is the more appropriate definition for the uncertainty region and on the desired degree of conservativeness.

For categorical missing data, the set $\boldsymbol{\Omega}$ may cover an in-depth grid of the whole parameter space of $\boldsymbol{\omega}$, but for continuous data, this strategy is clearly not feasible. Therefore, when there is no prior information to choose $\boldsymbol{\Omega}$ under a certain parameterization, the elicitation task may become easier if we reparameterize. For example, in lieu of using $\mu_{(0)}$ as sensitivity parameter, we may prefer to use α , β , or p , where $\mu_{(0)} = \alpha + \mu_{(1)}$, $\mu_{(0)} = \beta\mu_{(1)}$, chiefly for positive variables, and $\mu_{(0)} = F_{(1)}^{-1}(p)$, the p -th quantile of the theoretical distribution of the observed units. The variance of the estimator of μ depends upon which sensitivity parameter strategy is used, e.g., for the first three cases, we have

$$\text{Var} [\hat{\mu}(\mu_{(0)})] = \frac{\gamma_1 \sigma_{(1)}^2}{n} + \frac{\gamma_1(1 - \gamma_1)(\mu_{(1)} - \mu_{(0)})^2}{n}, \quad (3)$$

$$\text{Var} [\hat{\mu}(\alpha)] = \sigma_{(1)}^2 \text{E} \left(\frac{1}{n_1} \right) + \frac{\gamma_1(1 - \gamma_1)\alpha^2}{n}, \quad (4)$$

$$\text{Var} [\hat{\mu}(\beta)] = \sigma_{(1)}^2 \left[\beta^2 \text{E} \left(\frac{1}{n_1} \right) + \frac{2\beta(1 - \beta)}{n} + \frac{\gamma_1(1 - \beta)^2}{n} \right] + \frac{\gamma_1(1 - \gamma_1)\mu_{(1)}^2(1 - \beta)^2}{n}. \quad (5)$$

Since no sensible analysis can be accomplished if all outcomes are missing, i.e., if $n_1 = 0$, we could have computed (3)–(5) assuming that n_1 follows a positive binomial distribution (Stephan, 1945) instead of a binomial distribution with parameters n and γ_1 . However, the changes required for such purposes generate cumbersome formulae and improve accuracy only when $n\gamma_1$ is small. Simple approximations for the first negative moment of the positive

binomial distribution are discussed, for example, by Grab & Savage (1954) and Mendenhall & Lehman (1960). We adopt $E(1/n_1) \cong 1/(n\gamma_1+1-\gamma_1)$ in practice; this is usually accurate to at least two decimal places if $n\gamma_1 > 10$ (Grab & Savage, 1954). Nevertheless, comparing (3)–(5) is easier when using the cruder approximation $1/n\gamma_1$ and, then, the following relationships hold for large n :

$$\begin{aligned} \text{Var} [\hat{\mu}(\mu_{(0)})] &< \text{Var} [\hat{\mu}(\alpha)] &< \text{Var} [\hat{\mu}(\beta)], & \text{if } \beta \leq \beta_1 \text{ or } \beta \geq 1, \\ \text{Var} [\hat{\mu}(\mu_{(0)})] &\leq \text{Var} [\hat{\mu}(\beta)] &< \text{Var} [\hat{\mu}(\alpha)], & \text{if } \beta_1 < \beta \leq \beta_2 \text{ or } \beta_3 \leq \beta < 1, \\ \text{Var} [\hat{\mu}(\beta)] &< \text{Var} [\hat{\mu}(\mu_{(0)})] &< \text{Var} [\hat{\mu}(\alpha)], & \text{if } \beta_2 < \beta < \beta_3, \end{aligned} \quad (6)$$

where $\beta_1 = (-1 - \gamma_1)/(1 - \gamma_1)$, $\beta_2 = (-\gamma_1^4 - \gamma_1)/(1 - \gamma_1)$ and $\beta_3 = (\gamma_1^4 - \gamma_1)/(1 - \gamma_1)$. As $\beta_1 < \beta_2 < \beta_3 < 0$ and β will generally be positive, the last inequality of (6) will hardly occur in practice.

Using expressions for the asymptotic variances and covariances of order statistics (Sen, Singer & Pedroso de Lima, 2009, p.223), we obtain

$$\begin{aligned} \text{Var} [\hat{\mu}(p)] &\cong \frac{\gamma_1 \sigma_{(1)}^2}{n} + \frac{p(1-p)}{\{f_{(1)}[F_{(1)}^{-1}(p)]\}^2} \left[\frac{\gamma_1 - 2}{n} + E \left(\frac{1}{n_1} \right) \right] + \frac{\gamma_1(1-\gamma_1)[\mu_{(1)} - F_{(1)}^{-1}(p)]^2}{n} \\ &+ \frac{2}{n} \left[E \left(\frac{1}{n_1} \right) - \frac{1}{n} \right] \left\{ \sum_{j=1}^k \frac{p_j(1-p)}{f_{(1)}[F_{(1)}^{-1}(p_j)] f_{(1)}[F_{(1)}^{-1}(p)]} + \sum_{j=k+1}^{n_1} \frac{p(1-p_j)}{f_{(1)}[F_{(1)}^{-1}(p)] f_{(1)}[F_{(1)}^{-1}(p_j)]} \right\}, \end{aligned} \quad (7)$$

where $f_{(1)}$ denotes the density of the observed units, $p_k < p < p_{k+1}$ and $p_j = j/n_1 + o(n_1^{-1/2})$, $j = 1, \dots, n_1$, such that $0 < p_1 < p_2 < \dots < p_{n_1} < 1$, e.g., p_j may be equal to $(j - 0.5)/n_1$ or to one of the other three definitions discussed by Hyndman & Fan (1996) that satisfy their Property 5. Figure 1 portrays estimates of the square root of (3)–(5) and (7) for the data in Table 1, obtained by replacing the parameters in the formulae by their sample analogues; $f_{(1)}$ in (7) was replaced by a Gaussian kernel density estimate (Silverman, 1986). The 4 horizontal axes indicate equivalences among the 4 sensitivity parameters for estimating μ ;

for instance, $p = 0.9$, $\beta \cong 1.15$, $\alpha \cong 138$ and $\mu_0 \cong 1,083$ lead to the same $\hat{\mu}$ for public colleges, but to different estimates of its standard error.

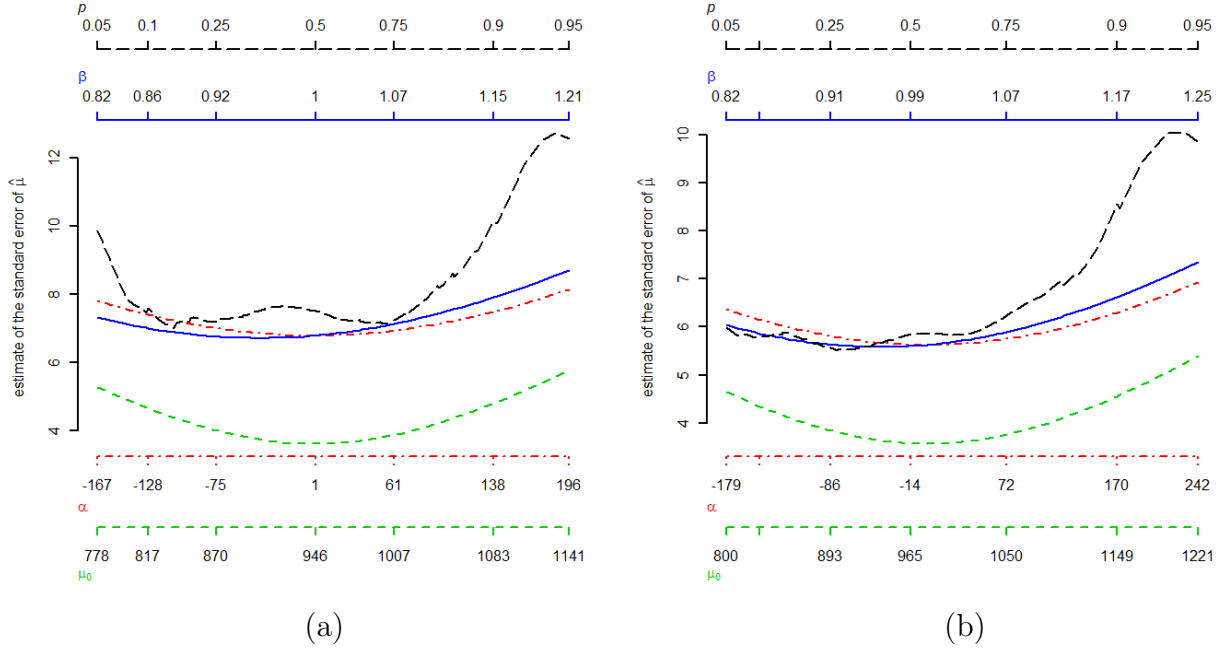


Figure 1: Estimates of standard errors of $\hat{\mu}$ using $\mu_{(0)}$, α , β or p as sensitivity parameter for (a) public and (b) private colleges

Assume that five experts were consulted and that their beliefs were interpreted according to the sensitivity parameter β , equally for public and private colleges, with $\Omega = [c; 1.00]$, where c was equal to 0.90, 0.91, 0.94, 0.95 and 0.96, respectively. That is, they stated that the mean CSAT for both public and private colleges that did not inform the CSAT is equal or slightly smaller than the mean CSAT for those colleges that reported it. Table 3 shows that the conclusions may be different depending on Ω and on the choice of the EURO version, which should be set a priori: the mean CSAT for public and private colleges is considered different for $\Omega = [0.96; 1.00]$ irrespectively of the adopted EURO, for $\Omega = [0.95; 1.00]$ when weak and pointwise EUROS are chosen, and for $\Omega = [0.94; 1.00]$ in the case of weak EURO; for the other combinations of Ω and EUROS, the results do not allow us to conclude whether there is a difference between the means. However, as the data of Section 2 are free from imprecision and all uncertainty comes from the ignorance regarding the missing data, HEIR

provides all the information we need.

Table 3: HEIR and 95% weak, pointwise and strong EUROs for the difference of the mean CSAT between private and public colleges

Ω for β	HEIR	Weak EURO	Pointwise EURO	Strong EURO
[0.90; 1.00]	[-2.3; 77.5]	[-5.9; 81.2]	[-16.8; 92.1]	[-19.6; 94.9]
[0.91; 1.00]	[1.3; 73.1]	[-2.9; 77.4]	[-13.2; 87.6]	[-15.0; 90.4]
[0.94; 1.00]	[12.0; 59.9]	[5.6; 66.3]	[-2.4; 74.4]	[-5.2; 77.1]
[0.95; 1.00]	[15.6; 55.5]	[8.3; 62.8]	[1.2; 69.9]	[-1.6; 72.7]
[0.96; 1.00]	[19.2; 51.1]	[10.8; 59.5]	[4.7; 65.5]	[2.0; 68.3]

Looking at the estimation of σ in (2), it may be more meaningful to work with λ in $\sigma_{(0)}^2 = \lambda^2 \sigma_{(1)}^2$ than with $\sigma_{(0)}$. A simple way to obtain an estimate of the variance of $\hat{\sigma}$, whether we use $(\mu_{(0)}, \sigma_{(0)}^2)$ or any other parameterization, is to employ the nonparametric bootstrap (Efron & Gong, 1983). As an example of how the standard error may vary considerably depending on the parameterization, setting $\sigma_0 = 134.3$ or $\lambda = 1.25$ for public colleges, both with $\beta = 0.90$, we obtain the same estimate for σ (129.6), but the estimates of the standard errors of $\hat{\sigma}$ are 2.3 and 5.1, respectively. In Table 4 we display estimated intervals for the mean and standard deviation of CSAT for each type of college administration using $\Omega = [0.90; 1.00]$ for β and $\Omega = [0.80; 1.25]$ for λ . Wherever the bootstrap has been employed in the paper, we generated 10,000 replicates of the statistic. In addition to what was previously concluded based on Table 3, the results of Table 4 do not suggest a difference in dispersion of the CSAT scores between public and private colleges.

Table 4: HEIR and 95% weak, pointwise and strong EUROs for μ and σ of CSAT using $\Omega = [0.90; 1.00]$ for β and $\Omega = [0.80; 1.25]$ for λ

Admin.	Parameter	HEIR	Weak EURO	Pointwise EURO	Strong EURO
Public	μ	[901.2; 945.3]	[897.0; 949.5]	[890.0; 956.4]	[887.8; 958.6]
	σ	[98.0; 129.6]	[95.3; 132.8]	[90.7; 138.0]	[89.3; 139.7]
Private	μ	[943.0; 978.8]	[939.3; 982.4]	[933.7; 988.0]	[931.9; 989.8]
	σ	[120.4; 149.5]	[117.8; 152.6]	[113.5; 157.5]	[112.1; 159.0]

4 Multivariate case

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})'$ where Y_{ij} denote the j -th response of the i -th unit of the study, $i = 1, \dots, n$, $j = 1, \dots, J$. In addition, define a vector of response indicators $\mathbf{R}_i = (R_{i1}, \dots, R_{iJ})'$ with $R_{ij} = 1$ if Y_{ij} is observed and $R_{ij} = 0$ otherwise. Relations (1) and (2) are then the univariate versions of

$$\boldsymbol{\mu} = \sum_{\mathbf{r} \in \mathcal{R}} \gamma_{\mathbf{r}} \boldsymbol{\mu}_{(\mathbf{r})}, \quad (8)$$

$$\boldsymbol{\Sigma} = \sum_{\mathbf{r} \in \mathcal{R}} \gamma_{\mathbf{r}} \boldsymbol{\Sigma}_{(\mathbf{r})} + \sum_{\mathbf{r} \in \mathcal{R}} \gamma_{\mathbf{r}} (\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})(\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})', \quad (9)$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_J)' = \text{E}(\mathbf{Y}_i)$, $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{Y}_i)$, \mathcal{R} contains all observed values of \mathbf{r} , $\gamma_{\mathbf{r}} = P(\mathbf{R}_i = \mathbf{r})$, $\boldsymbol{\mu}_{(\mathbf{r})} = (\mu_{1(\mathbf{r})}, \dots, \mu_{J(\mathbf{r})})' = \text{E}(\mathbf{Y}_i | \mathbf{R}_i = \mathbf{r})$ and $\boldsymbol{\Sigma}_{(\mathbf{r})} = \text{Cov}(\mathbf{Y}_i | \mathbf{R}_i = \mathbf{r})$. Often, we prefer to work with correlations $\psi_{jk(\mathbf{r})} = \text{Corr}(Y_{ij}, Y_{ik} | \mathbf{R}_i = \mathbf{r})$ instead of covariances $\sigma_{jk(\mathbf{r})} = \text{Cov}(Y_{ij}, Y_{ik} | \mathbf{R}_i = \mathbf{r})$, $j \neq k$, and therefore we let $\boldsymbol{\Sigma}_{(\mathbf{r})} = \mathbf{D}_{\boldsymbol{\sigma}_{(\mathbf{r})}} \boldsymbol{\Psi}_{(\mathbf{r})} \mathbf{D}_{\boldsymbol{\sigma}_{(\mathbf{r})}}$, where $\mathbf{D}_{\boldsymbol{\sigma}_{(\mathbf{r})}}$ denotes a diagonal matrix with the elements of $\boldsymbol{\sigma}_{(\mathbf{r})}$ along the main diagonal, $\boldsymbol{\sigma}_{(\mathbf{r})} = (\sigma_{j(\mathbf{r})}, j = 1, \dots, J)'$, $\sigma_{j(\mathbf{r})}^2 = \text{Var}(Y_{ij} | \mathbf{R}_i = \mathbf{r})$ and $\boldsymbol{\Psi}_{(\mathbf{r})} = \text{Corr}(\mathbf{Y}_i | \mathbf{R}_i = \mathbf{r})$; the corresponding definitions for the unconditional variances, covariances and correlations follow analogously.

In the bivariate version of Table 2, there are 4 missingness patterns and 11 non-identifiable parameters for estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using (8) and (9). They are 4 means, 4 standard deviations and 3 correlations, as indicated by the non-observed values in Table 2. For $J = 10$ variables subject to missingness, there are $2^J = 1,024$ potential missingness patterns that would lead to 44,800 non-identifiable parameters: $J \times 2^{J-1} = 5,120$ means, 5,120 standard deviations and $\sum_{j=0}^{J-1} \binom{J}{j} \left[\binom{J}{2} - \binom{j}{2} \right] = 34,560$ correlations. The challenge here is not only that the number of sensitivity parameters may increase exponentially depending on the missingness patterns and number of variables, but also that there are additional options for the parameterization. For instance, in the bivariate problem, instead of using $\mu_{1(0,1)}$ and $\mu_{1(0,0)}$

as sensitivity parameters, we may prefer to employ functions of these parameters along with the identifiable ones, namely, $\mu_{1(1,1)}$ and $\mu_{1(1,0)}$, or yet to use other alternatives, such as quantiles of the distribution of Y_{i1} conditional on $\mathbf{R}_i = (1, 1)$ or on $\mathbf{R}_i = (1, 0)$. Optionally, we may decide to use the distribution of Y_{i1} conditional on $R_{i1} = 1$, that is, all available data for this variable, or its mean $\mu_{1(1)} = E(Y_{i1}|R_{i1} = 1)$ as, for example,

$$\mu_{1(0,1)} = \beta_{11}\mu_{1(1)}, \quad \mu_{1(0,0)} = \beta_{10}\mu_{1(1)}.$$

Likewise, we may reparameterize

$$\begin{aligned} \mu_{2(1,0)} &= \beta_{21}\mu_{2(1)}, & \mu_{2(0,0)} &= \beta_{20}\mu_{2(1)}, \\ \sigma_{1(0,1)}^2 &= \lambda_{11}^2\sigma_{1(1)}^2, & \sigma_{1(0,0)}^2 &= \lambda_{10}^2\sigma_{1(1)}^2, \\ \sigma_{2(1,0)}^2 &= \lambda_{21}^2\sigma_{2(1)}^2, & \sigma_{2(0,0)}^2 &= \lambda_{20}^2\sigma_{2(1)}^2, \end{aligned}$$

where $\mu_{2(1)} = E(Y_{i2}|R_{i2} = 1)$ and $\sigma_{j(1)}^2 = \text{Var}(Y_{ij}|R_{ij} = 1)$, $j = 1, 2$.

Because of the large number of alternative parameterizations that may be combined for a set of variables, we computed the variance of $\hat{\boldsymbol{\mu}}$ only for the case where the sensitivity parameters are the non-identifiable means in $\boldsymbol{\mu}_{(\mathbf{r})}$. In the Appendix, this result as well as (8) and (9) are expressed in matrix formulation useful for computational implementation. Again we appeal to the non-parametric bootstrap to obtain estimates for the standard error of $\hat{\boldsymbol{\mu}}$ under other parameterizations for the sensitivity parameters as well as of $\hat{\boldsymbol{\sigma}}$, $\hat{\boldsymbol{\Psi}}$ and functions thereof.

In Table 5 we exhibit estimated intervals for means and standard deviations of GRADRAT and CSAT and their correlation using $\boldsymbol{\Omega} = [0.90; 1.00]$ for $\{\beta_{jk}\}$, $\boldsymbol{\Omega} = [0.80; 1.25]$ for $\{\lambda_{jk}\}$ and $\boldsymbol{\Omega} = [0; 1]$ for $\psi_{12(\mathbf{r})}$, $\mathbf{r} \neq (1, 1)$. The lower and upper bounds of the ignorance region of each parameter were obtained, respectively, by minimizing and maximizing the corresponding function over the sensitivity parameters. Note, though, that all 11 sensitivity parameters

need to be used only when the target is the correlation; as presumed by (1) and (2), when the target is a mean of one variable, the optimizations may be carried over the non-identifiable means of this variable (i.e., 2 sensitivity parameters) and, for standard deviations, it may be performed over the non-identifiable means and standard deviations (4 sensitivity parameters). From the results displayed in Table 5, we conclude that GRADRAT and CSAT are positively linearly correlated; the magnitude of the correlation, however, is difficult to assess, given the ignorance caused by the missing data. This conclusion remains when only the set Ω corresponding to $\psi_{12(\mathbf{r})}$, $\mathbf{r} \neq (1, 1)$, is changed to $[-0.5; 1]$. This notwithstanding, if we believe that GRADRAT and CSAT may have a linear correlation smaller than -0.5 when any of the variables is missing, the results do not allow us to conclude whether the two variables are linearly correlated.

Table 5: HEIR and 95% weak, pointwise and strong EUROs for μ , σ and ψ_{12} of GRADRAT and CSAT using $\Omega = [0.90; 1.00]$ for $\{\beta_{jk}\}$, $\Omega = [0.80; 1.25]$ for $\{\lambda_{jk}\}$ and $\Omega = [0; 1]$ for $\psi_{12(\mathbf{r})}$, $\mathbf{r} \neq (1, 1)$

Interval	GRADRAT		CSAT		Correlation
	Mean	SD	Mean	SD	
HEIR	[59.9; 60.4]	[18.6; 19.3]	[929.1; 968.0]	[114.3; 144.8]	[0.288; 0.795]
Weak EURO	[59.1; 61.2]	[18.2; 19.7]	[927.4; 969.7]	[112.8; 146.6]	[0.300; 0.786]
Pointwise EURO	[59.0; 61.3]	[18.0; 19.8]	[922.3; 975.1]	[108.8; 151.1]	[0.262; 0.817]
Strong EURO	[58.9; 61.4]	[17.9; 20.0]	[921.0; 976.5]	[107.7; 152.3]	[0.257; 0.821]

5 Concluding remarks

Selection models and pattern-mixture models are likely the most common frameworks for incomplete data modelling. In the univariate case, Scharfstein, Daniels & Robins (2003) showed that the assumption

$$\text{logit } P(R_i = 0|Y_i = y) = \text{constant} + q(y) \tag{10}$$

under the selection model is equivalent to the restriction

$$f_{(0)}(y) = f_{(1)}(y) \frac{\exp[q(y)]}{\int_{-\infty}^{\infty} \exp[q(s)] f_{(1)}(s) ds}, \quad \forall y, \quad (11)$$

under the pattern-mixture model, where q is the so-called selection-bias function. Beyond the direct interpretation of these expressions, they also noted that, if $q(y) = \delta \log(y)$, for example, we obtain from (10) that $\exp(\delta)$ is the odds ratio of missingness between subjects who differ by one unit of $\log(y)$; from (11), $\delta > 0$ (< 0) indicates that the distribution of Y_i for the missing outcomes is more (less) heavily weighted towards large values of Y_i than the distribution of Y_i for the observed outcomes. These insights are fundamental to carry out a sensitivity analysis. Nevertheless, the functional form of $q(y)$ as well as the range of values to be considered for δ are hard to assess.

When the target of inference is the mean, the standard deviation, the correlation or some function thereof, we may employ their non-identifiable counterparts as sensitivity parameters. Elicitation is never a trivial task, but these sensitivity parameters are easier to elicit than the selection-bias functions because the former are directly related to the parameters of interest. However, there are connections between both strategies; for example, for a specified $q(y)$, we can use (11) to compute the corresponding results for $\mu_{(0)}$, α , β , p , $\sigma_{(0)}$, and λ of Section 3.

Some of these ideas on parameterization were considered previously in the literature. For instance, Rubin (1977) uses (1) to develop a Bayesian solution assuming normality and Daniels & Hogan (2000) consider a pattern-mixture model of multivariate normal distributions wherein the model identification is accomplished through $\mathbf{b}_{(d)} = \boldsymbol{\mu}_{(d)} - \boldsymbol{\mu}_{(d+1)}$ and $\mathbf{C}_{(d)} = \boldsymbol{\Sigma}_{(d)}^{1/2} \boldsymbol{\Sigma}_{(d+1)}^{-1/2}$, $d = 1, \dots, J$, where $d = 1 + \sum_{j=1}^J r_j$ is the drop-out indicator and $\mathbf{b}_{(d)}$ and $\mathbf{C}_{(d)}$ are, respectively, pre-specified vectors and matrices. We extended the results by dropping the normality assumption and allowing a greater flexibility for the identification of the model. First, we not only considered absolute differences of means of missingness patterns but also relative differences and the possibility of relating non-observed means to

quantiles of observed distributions. Second, we replaced the hard-to-elicited functions of covariance matrices by relative differences of standard deviations and by correlations. Third, in the multivariate case, we showed that a non-identifiable mean (or standard deviation) may be related to an identifiable one not only in cases with a single specific missingness pattern but also in cases with sets of missingness patterns wherein the corresponding variable is observed. With these alternatives, it is easier to extract information from experts or historical data and, consequently, to produce more meaningful and more plausible sensitivity analyses.

When Vansteelandt et al. (2006) show how to construct EUROs, they assume that the values of the sensitivity parameters that correspond to the lower and upper bounds of the ignorance interval are independent of the observed data. This assumption is satisfied when the target of inference is the mean, but it may fail for standard deviations and correlations. Consider, for example, the target σ under the univariate case of Section 3, and assume that $\mathbf{\Omega}_{\sigma_{(0)}} = [\sigma_{(0)}^L; \sigma_{(0)}^U]$ and $\mathbf{\Omega}_{\mu_{(0)}} = [\mu_{(0)}^L; \mu_{(0)}^U]$ are specified. Looking at (2) after replacing the identifiable parameters by their sample counterparts, we note that $\sigma_{(0)}^L$ and $\sigma_{(0)}^U$ are the values of $\sigma_{(0)}$ in the set $\mathbf{\Omega}_{\sigma_{(0)}}$ that, respectively, minimize and maximize $\hat{\sigma}(\mu_{(0)}, \sigma_{(0)})$ irrespectively of the data and of $\mu_{(0)}$. However, as

$$\arg \min_{\mu_{(0)} \in \mathbf{\Omega}_{\mu_{(0)}}} \hat{\sigma}(\mu_{(0)}, \sigma_{(0)}) = \begin{cases} \mu_{(0)}^L, & \text{if } \hat{\mu}_{(1)} < \mu_{(0)}^L, \\ \mu_{(0)}^U, & \text{if } \hat{\mu}_{(1)} > \mu_{(0)}^U, \\ \hat{\mu}_{(1)}, & \text{otherwise,} \end{cases}$$

$$\arg \max_{\mu_{(0)} \in \mathbf{\Omega}_{\mu_{(0)}}} \hat{\sigma}(\mu_{(0)}, \sigma_{(0)}) = \begin{cases} \mu_{(0)}^U, & \text{if } (\hat{\mu}_{(1)} - \mu_{(0)}^U)^2 > (\hat{\mu}_{(1)} - \mu_{(0)}^L)^2, \\ \mu_{(0)}^L, & \text{otherwise} \end{cases}$$

clearly depends on the data through $\hat{\mu}_{(1)}$, the assumption is violated. We note though that, for this specific case with only two missingness patterns, the assumption is satisfied if we switch the sensitivity parameter $\mu_{(0)}$ to α or β . Vansteelandt et al. (2006) present another example wherein the assumption fails and, therefore, compute bootstrap estimates of the

coverage probability to show that no undercoverage or overcoverage is detected. Nevertheless, in a set-up like ours, where the estimate of the standard error is already computed with the bootstrap, implications for computation time would be considerable.

Acknowledgements

We gratefully acknowledge the financial supports to this research: Frederico Z. Poletto 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; Geert Molenberghs, from the IAP research Network P6/03 of the Belgian Government (Belgian Science Policy); Carlos Daniel Paulino, from Fundação para a Ciência e Tecnologia (FCT) through the research centre CEAUL-FCUL, Portugal.

References

- ALLISON, P. D. (2001). *Missing Data*. Thousand Oaks: Sage.
- DANIELS, M. J. & HOGAN, J. W. (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics* **56**, 1241–1248.
- EFRON, B. & GONG, G. (1983). A leisure look at the bootstrap, the jackknife and cross-validation. *Amer. Statist.* **37**, 36–48.
- FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. & MOLENBERGHS, G. (2008). *Longitudinal Data Analysis*. Boca Raton: Chapman & Hall / CRC.
- GLYNN, R. J., LAIRD, N. M. & RUBIN, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse (with discussion). In H. Wainer, ed., *Drawing Inferences from Self-Selected Samples*. Mahwah: Lawrence Erlbaum Associates, 115–151.

- GRAB, E. L. & SAVAGE, I. R. (1954). Tables of the expected value of $1/X$ for positive Bernoulli and Poisson variables. *J. Amer. Statist. Assoc.* **49**, 169–177.
- HYNDMAN, R. J. & FAN, Y. (1996). Sample quantiles in statistical packages. *Amer. Statist.* **50**, 361–365.
- LITTLE, R. J. A. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika* **81**, 471–483.
- LITTLE, R. J. A. & RUBIN, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed. New York: John Wiley & Sons.
- LITTLE, R. J. A. & WANG, Y.-X. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* **52**, 98–111.
- MENDENHALL, W. & LEHMAN, J., E. H. (1960). An approximation to the negative moments of the positive binomial useful in life testing. *Technometrics* **2**, 227–242.
- ROTNITZKY, A., ROBINS, J. M. & SCHARFSTEIN, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J. Amer. Statist. Assoc.* **93**, 1321–1339.
- ROTNITZKY, A., SCHARFSTEIN, D., SU, T.-L. & ROBINS, J. M. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics* **57**, 103–113.
- RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- RUBIN, D. B. (1977). Formalizing subjective notions about the effect of nonrespondents in sample surveys. *J. Amer. Statist. Assoc.* **72**, 538–543.
- 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). *J. Amer. Statist. Assoc.* **94**, 1096–1146.
- SEN, P. K., SINGER, J. M. & PEDROSO DE LIMA, A. C. (2009). *From Finite Sample to Asymptotic Methods in Statistics*. Cambridge: Cambridge University Press.
- SILVERMAN, B. W. (1986). *Density Estimation*, 2nd ed. London: Chapman & Hall.
- STEPHAN, F. F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *Ann. Math. Statist.* **16**, 50–61.
- VANSTEELENDT, 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.

Appendix

With some algebra, (8) and (9) may be conveniently rewritten as

$$\begin{aligned}\boldsymbol{\mu} &= \boldsymbol{\gamma}'_* \boldsymbol{\mu}_{(*)}, \\ \boldsymbol{\Sigma} &= \boldsymbol{\gamma}'_* \boldsymbol{\Sigma}_{(*)} (\mathbf{1}_P \otimes \mathbf{I}_J) + \boldsymbol{\gamma}'_* \mathbf{S}_{(*)} (\mathbf{1}_P \otimes \mathbf{I}_J),\end{aligned}$$

where $\boldsymbol{\gamma}_* = \boldsymbol{\gamma} \otimes \mathbf{I}_J$, $\boldsymbol{\gamma} = (\gamma_{\mathbf{r}}, \mathbf{r} \in \mathcal{R})'$, \otimes denotes the Kronecker product, \mathbf{I}_J represents an identity matrix of order J , $\boldsymbol{\mu}_{(*)} = (\boldsymbol{\mu}'_{\mathbf{r}}, \mathbf{r} \in \mathcal{R})'$, $\boldsymbol{\Sigma}_{(*)}$ and $\mathbf{S}_{(*)}$ are block diagonal matrices with blocks $\boldsymbol{\Sigma}_{(\mathbf{r})}$ and $\mathbf{S}_{(\mathbf{r})} = (\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})(\boldsymbol{\mu}_{(\mathbf{r})} - \boldsymbol{\mu})'$, $\mathbf{r} \in \mathcal{R}$, respectively, $\mathbf{1}_P$ denotes a $P \times 1$ vector with all elements equal to 1 and P represents the number of missingness patterns, i.e., the cardinality of \mathcal{R} . When employing as sensitivity parameters the non-identifiable means

of $\boldsymbol{\mu}_{(\mathbf{r})}$, stacked in the vector $\boldsymbol{\mu}^{\text{NI}}$, the covariance matrix of $\hat{\boldsymbol{\mu}}$ is specified as

$$\text{Cov} [\hat{\boldsymbol{\mu}}(\boldsymbol{\mu}^{\text{NI}})] = \frac{1}{n} \boldsymbol{\gamma}'_* \boldsymbol{\Sigma}_{(*)}^I (\mathbf{1}_P \otimes \mathbf{I}_J) + \frac{1}{n} \boldsymbol{\mu}'_D [(\mathbf{D}_\gamma - \boldsymbol{\gamma} \boldsymbol{\gamma}') \otimes (\mathbf{1}_J \mathbf{1}'_J)] \boldsymbol{\mu}_D,$$

where $\boldsymbol{\mu}_D = (\mathbf{D}_{\boldsymbol{\mu}_{\mathbf{r}}}, \mathbf{r} \in \mathcal{R})'$ and $\boldsymbol{\Sigma}_{(*)}^I$ is obtained from $\boldsymbol{\Sigma}_{(*)}$ by replacing the non-identifiable parameters by 0.