Spatial small area smoothing models for handling survey data with nonresponse

Peer-reviewed author version

WATJOU, Kevin; FAES, Christel; LAWSON, Andrew; Kirby, R. S.; AREGAY, Mehreteab; Carroll, R. & VANDENDIJCK, Yannick (2017) Spatial small area smoothing models for handling survey data with nonresponse. In: STATISTICS IN MEDICINE, 36(23), p. 3708-3745.

DOI: 10.1002/sim.7369

Handle: http://hdl.handle.net/1942/24962

Spatial small area smoothing models for handling survey data with nonresponse

K. Watjou[1], C. Faes[1], A. Lawson[2], R.S. Kirby[3], M. Aregay[2], R. Carroll[2], and Y. Vandendijck[1]

[1] *Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium*
[2] *Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, USA*
[3] *Department of Community and Family Health, College of Public Health, University of South Florida, Tampa, FL, USA*

Abstract

Spatial smoothing models play an important role in the field of small area estimation (SAE). In the context of complex survey designs, the use of design weights is indispensable in the estimation process. Recently, efforts have been made in these spatial smoothing models, in order to obtain reliable estimates of the spatial trend. However, the concept of missing data remains a prevalent problem in the context of spatial trend estimation as estimates are potentially subject to bias. In this paper, we focus on spatial health surveys where the available information consists of a binary response and its associated design weight. Furthermore, we investigate the impact of nonresponse as missing data on a range of spatial models for different missingness mechanisms and different degrees of missingness by means of an extensive simulation study. The computations were done in R, using INLA and other existing packages. The results show that weight adjustment to correct for missingness has a beneficial effect on the bias in the missing at random (MAR) setting for all models. Furthermore we estimate the geographical distribution of perceived health at the district level based on the Belgian Health Interview Survey (2001).

Keywords: Complex Survey Design, Disease Mapping, Hierarchical Bayesian Modeling, Integrated Nested Laplace Approximation, Missing Data.

1 Introduction

Health surveys are an important source of information when investigating the geographical distribution of diseases. Hierarchical spatial smoothing methods are well developed and used in a fairly standard manner in disease mapping to infer about the geographical distribution of diseases (see e.g. Elliott *et al.* [1], Waller and Gotway [2], Lawson [3]). However, health surveys are typically complex in design, with complex sample selection methods for drawing the sample from the population. Examples of possible sampling designs include stratified sampling, cluster sampling, convenience sampling, etc. An overview can be found in Schaeffer *et al.* [4].

Approaches to take into account the complex survey design can be grouped into design-based, area-level and unit-level approaches. The most commonly used design-based method is the Horvitz-Thompson (HT) estimator [5]. The HT estimator is a weighted estimator, with weights equal to the inverse of the sampling probability. The idea of weighting is to make the sample as similar as possible to the population. This estimator is design consistent and provides reliable inferences in large samples, but it can be very inefficient, especially when sample sizes are small (Basu [6], Rao [7]). In unit-level approaches, inference is built on a model of the health outcome, which takes into account all key features of the sampling design such as weighting, stratification and clustering. These models can become very complex as a large number of variables need to be included in the model. In addition, key variables that are required for inclusion of individuals in the sample may be unavailable (Gelman [8], Little [9], Pfefferman [10]). Design-based inference can be model-assistent, such as the generalized regression estimator (GREG), bringing together model features and design-based inference. Chen *et al.* [11], Mercer *et al.*[12] and Vandendijck *et al.* [13] describe methods of incorporating the design weights within a spatial hierarchical model. In this paper, we compare a number of these proposed methods and investigate the effect of nonresponse on the estimation of the geographical distribution of the outcome of interest.

In addition, nonresponse often occurs in surveys. It is often the case that selected individuals may not want to participate in the survey or answer only part of the questions. The effect of this missingness on inference is two-fold. First, as it reduces the sample size, it will lead to a reduced precision of the estimates. Second, incompleteness can lead to biased inference, as the population that do respond to the question might differ systematically from the population that do not respond to the question of interest. Different analysis approaches exist for dealing with incompleteness of data. Focusing on incomplete data under the missing at random (MAR) assumption, we can group the methods into weighting methods, imputation methods and full information maximum likelihood methods (Rubin [14], Little *et al.* [15]). In the weighing methods, subjects that have no missing observations are weighted in order to compensate for the removal of subjects that do have missing observations. In the imputation methods, plausible values for the missing observations are filled in. Full information maximum likelihood methods using only the available data yield appropriate likelihood-based inference under an MAR mechanism. Weighting has been widely used in many public health studies, and therefore is the focus of interest in this paper.

The goal of this paper is to investigate the effect of nonresponse in a complex survey on the estimated geographical distribution of disease prevalence and compare the performance of different models under missingness. In Section 2, we revise spatial smoothing methods that can be used to estimate the geographical distribution of the health outcome in the presence of a complex sampling design. Also, weight adjustments are discussed for incomplete data. In Section 3, a simulation study is conducted to investigate the impact of missingness of the estimated spatial distribution. Section 4 presents an application of the methods to investigate geographical differences of individual's perceived health based on the Belgian Health Interview Survey, with emphasis on the complex sampling design and incompleteness of the outcome.

2 Methodology

Denote Y_{ik} as the binary response value of the i^{th} individual in area k ($i = 1, \ldots, N_k$ and $k = 1, \ldots, K$), with N_k the population size in area k and $N = \sum_{k=1}^{K} N_k$ the overall population size. We assume that the population size is known for each area k. Interest is in the area-specific population prevalence P_k, defined as

$$P_k = \frac{1}{N_k} \sum_{i=1}^{N_k} Y_{ik}. \tag{1}$$

In order to get an estimate of the area-specific prevalence, a random probability-sample is taken from the population, in which individuals are sampled from the population with a known sample probability π_{ik}. The sample size in area k is n_k, where some of the n_k could be zero. We define $n = \sum_{k=1}^{K} n_k$ as the total sample size. The sampled responses are denoted by y_{ik}. Note that some of these responses could be missing, as some of the individuals might not respond to the question. We define an indicator variable r_{ik} which indicates whether the sampled individual i in area k responded to the question ($r_{ik} = 1$) or not ($r_{ik} = 0$). We define s_k as the total set of individuals which are sampled from district k, where $|s_k| = n_k$, and s_k^* as the set of individuals that responded to the question of interest, where $|s_k^*| = m_k$ represents the number of individuals which responded in area k.

The area-specific unweighted mean estimator based on the available data can be expressed as

$$\hat{P}_k^{UNW} = \frac{1}{m_k} \sum_{i \in s_k^*} y_{ik}. \tag{2}$$

While this estimator is unbiased in the situation of a simple random sampling design without replacement and missingness occurring completely at random, it lacks the ability to take the design and missingness features into account due to the absence of the sampling weights in the estimation process. In the next sections, different approaches are discussed that can take into account both the sampling design and incompleteness. Sections 2.1 and 2.2. give an overview of available methods to account for the non-random sampling design. Section 2.3 then discusses how these methods can be adapted to account for incompleteness of the sample.

2.1 Horvitz-Thompson Estimator

Weighting is commonly used in the analysis of survey data. As surveys are often characterised by a complex design, statistical methods need to take into account the design in order to correct for the loss of representation of the population. The idea of weighting is to make the sample as similar as possible to the population, by assigning a weight to each individual in the sample. This design weight could be calculated as the reciprocal of the probability of being sampled, namely $w_{ik}^d = \frac{1}{\pi_{ik}}$.

The famous Horvitz-Thompson estimator for the area-specific prevalence is

$$\hat{P}_k^{HT} = \frac{\sum_{i \in s_k} w_{ik}^d y_{ik}}{\sum_{i \in s_k} w_{ik}^d} = \frac{1}{n_k} \sum_{i \in s_k} \tilde{w}_{ik}^d y_{ik} \tag{3}$$

(Horvitz and Thompson [5]), with $\tilde{w}_{ik}^d$ the normalised design weight defined as

$$\tilde{w}_{ik}^d = n_k \cdot \frac{w_{ik}^d}{\sum_{i \in s_k} w_{ik}^d}. \tag{4}$$

This normalisation involves the reweighting of the sample to match the sample size in area k, i.e. $\sum_{i \in s_k} \tilde{w}_{ik}^d = n_k$. The variance of $\hat{P}^{HT}$ can be expressed as follows

$$\widehat{var}(\hat{P}_k^{HT}) = \frac{1}{n_k}\left(1 - \frac{n_k}{N_k}\right)\frac{1}{n_k - 1}\sum_{i \in s_k} \tilde{w}_{ik}^{d^2}(y_{ik} - \hat{P}_k^{HT})^2. \tag{5}$$

The Horvitz-Thompson estimator is a design-unbiased estimator of P_k. Note that this direct estimator uses only the observation from the area of interest (Rao [16]). However, when sample sizes are too small to produce reliable or stable estimates, it is better to use a unit-level estimator that borrows strength across the different areas by using the observations from all sampled individuals.

2.2 Area-Level Methods

Unlike the design-based methods, the area-level approaches assume a model conditional on the sampled observations and generally provide more accurate estimates (Pfeffermann [17]). The estimates of an area obtained by a design-based approach are considered to be direct as they are based solely on the measurements of the given geographical unit. Area-Level methods on the other hand produce indirect estimates, as these methods rely on the presumption that area-specific estimates borrow information from other areas as well. This makes it possible to find more accurate estimates (see e.g. Ugarte *et al.* [18], Salvati *et al.* [19], Chambers *et al.* [20], Rao and Molina [21]). Also, it creates the advantage that estimates can be obtained in areas with no sample, as opposed to the design-based methods where the observations within an area are assumed to be independent of observations acquired from surrounding areas. A linear model with area-specific random effects was first proposed by Fay and Herriot [22] in order to obtain survey estimates for income. Numerous applications have originated from the Fay-Herriot model, some of which are included in the following section.

In this section, we present several Bayesian hierarchical smoothing models that were used by Vandendijck *et al.* [13] and Mercer *et al.* [12], each consisting of three stages. At the first stage, the likelihood of the response is defined conditional on some latent variables (random effects). At the second stage, the latent variables are defined, whether or not on a transformed scale. At the third stage, the prior distributions on the variance parameters for the random effects and on any other unknown parameters are specified. In all models, vague priors were specified in order to minimize their effect on the inferential evaluation.

2.2.1 Unadjusted Binomial Model (UB)

The simplest approach ignores the sampling design. At the first stage it is assumed that

$$\begin{aligned} y_k|P_k &\sim \text{Binomial}(n_k, P_k) \\ \text{logit}(P_k) &= \beta_0 + u_k + v_k, \end{aligned} \tag{6}$$

where $y_k = \sum_{i \in s_k} y_{ik}$ are the aggregated responses in area k and u_k and v_k are area-specific random effects. At the second stage, we assumed a normal distribution for the uncorrelated heterogeneity which describes the heterogeneity in the data , i.e., $v_k \sim N(0, \sigma_v^2)$, whereas we considered an intrinsic conditional autoregressive model (ICAR) (Besag *et al.* [23], Rue *et al.* [24]) for the correlated heterogeneity u_k as follows:

$$u_k|u_{k',k \neq k'} \sim N\left(\frac{1}{a_k}\sum_{k' \in ne(k)} u_k^{'}, \frac{\sigma_u^2}{a_k}\right), \tag{7}$$

where $ne(k)$ indicates the set of neighbors and a_k the number of neighbors for a given area k. According to common convention, two areas are considered neighbors if they share a common boundary (Lawson [25]). Other choices for the neighborhood scheme are discussed by Bivand *et al.* [26].

Compared to the Horvitz-Thompson estimator, this model allows to account for both spatial dependence and heterogeneity, via the random effects specifications. This will result in the smoothing of extreme local estimates in areas with small sample sizes, which is a desirable effect as it prevents over-fitting. However, this model is not adjusted for the survey design, as the sample weights are not included in the estimation process. This implies that if the design of the survey is informative, the resulting estimates of the Unadjusted Binomial will be rendered biased. The models described below resolve this by actively using the sample weights to adjust for the outcome of interest y_k, while retaining the advantage of the reduction in variability.

2.2.2 Logit Normal Model (LN)

A simple way to allow spatial smoothing of the Horvitz-Thompson estimator is to assume a hierarchical spatial smoothing model for $\hat{P}_k^{HT}$. Because $\hat{P}_k^{HT}$ is restricted between 0 and 1, an empirical logit transformation of $\hat{P}_k^{HT}$ is considered. The resulting model is given by:

$$\begin{aligned} y_k^{LN}|P_k &\sim N\left(\text{logit}(P_k), \sigma_k^2\right) \\ \text{logit}(P_k) &= \beta_0 + u_k + v_k, \end{aligned} \tag{8}$$

where $y_k^{LN} = \text{logit}(\hat{P}_k^{HT})$ and the variance σ_k^2 is set equal to $\widehat{\text{var}}(\hat{P}_k^{HT})/\left(\hat{P}_k^{2^{HT}}(1-\hat{P}_k^{HT})^2\right)$ (see Mercer *et al*. [12]).

2.2.3 Arcsine Root Normal Model (AN)

As an alternative to the LN model, Raghunathan *et al*. [27] proposed the use of an arcsine-square root transformation of the direct estimates. This method assures that the sampling variances, which usually depend on the population proportions, are stabilised approximately (see also Efron and Morris [28]). This leads to the following model specification:

$$\begin{aligned} y_k^{AN}|P_k &\sim N\left(\sin^{-1}\left(\sqrt{P_k}\right), \sigma_k^2\right) \\ \sin^{-1}\left(\sqrt{\hat{P}_k}\right) &= \beta_0 + u_k + v_k, \end{aligned} \tag{9}$$

where $y_k^{AN} = \sin^{-1}\left(\sqrt{\hat{P}_k^{HT}}\right)$ and the variability $\sigma_k^2 = \frac{1}{4 \cdot n_k^E}$ depends on the effective sample size $n_k^E = \hat{P}_k^{HT}(1-\hat{P}_k^{HT})/\widehat{\text{var}}\left(\hat{P}_k^{HT}\right)$.

2.2.4 Pseudo-Likelihood Model (PL)

The method described by Congdon and Lloyd [29] employs a weighted likelihood whereby the response values are weighted using the normalised design weights $\tilde{w}_{ik}^d$ (see also Asparouhov [30]). Mercer *et al*. [12] noted that this model can be re-written as a simple hierarchical model of the form

$$\begin{aligned} y_k^{PL}|P_k &\sim \text{Binomial}(n_k, P_k) \\ \text{logit}(P_k) &= \beta_0 + u_k + v_k, \end{aligned} \tag{10}$$

with $y_k^{PL} = \sum_{i \in s_k} \tilde{w}_{ik}^d y_{ik}$, where $\tilde{w}_{ik}^d$ is defined as in (4).

2.2.5 Effective Sample Size Method (ES)

Chen *et al.* [11] proposed a similar approach to the pseudo-likelihood model, but accounting for the sampling design via the use of the effective sample size (n^E). The effective sample is defined as the sample size that is needed to match the variance from a simple random sample with that under a complex sampling design. Under the assumption of a simple stratified sampling design, this leads to an effective sample size n^E equal to

$$n_k^E = \widehat{P}_k^{HT}(1-\widehat{P}_k^{HT}) \Big/ \widehat{\mathrm{Var}}(\widehat{P}_k^{HT})$$

with $\widehat{\mathrm{Var}}(\widehat{P}_k^{HT})$ as specified in Section 2.1. The model can be described as:

$$\begin{aligned} y_k^E | P_k &\sim \text{Binomial}(n_k^E, P_k) \\ \text{logit}(P_k) &= \beta_0 + u_k + v_k, \end{aligned} \tag{11}$$

where $y_k^E = n_k^E \cdot \hat{P}_k^{HT}$ represents the effective number of cases. The use of this adjusted binomial likelihood gives a better reflection of the sampling distribution as compared to the normal approximation in (8).

2.3 Unit-Level Methods (MB)

In contrast to previous area-level methods, Royall [31] proposed a predictive hierarchical model at the unit level, in order to define an estimator for P_k:

$$\hat{P}_k = \frac{1}{N_k}\left(\sum_{i\in s_k} y_{ik} + \sum_{i\in s_k'} \hat{y}_{ik}\right). \tag{12}$$

The first term sums up the observed response values of the sampled individuals in area k, whereas the second term refers to the unobserved individuals from the population and needs to be estimated from the sample. A flexible model is formulated for the observed data y_{ik}, which is then used to predict the response values for the non-sampled individuals ($\hat{y}_{ik}$). Note that, the prediction model for y_{ik} should take into account the variables that affected the sampling design, as the responses could depend on these characteristics. As this might lead to a prediction model with many covariates and as commonly not all covariates are publicly available, the available design weights can be used as a proxy for the population strata and used as a variable in the prediction model (Zheng and Little [32-33]).

Chen *et al.* [34] proposed a Bayesian penalised spline predictive estimator in a survey sampling setting for a finite population proportion, whereby the inclusion probability was incorporated directly into the model as a covariate by means of a binary p-spline probit regression model. Vandendijck *et al.* [13] extended these ideas to the context of small area estimation. Two versions of the hierarchical weight-smoothing model are considered:

$$\begin{aligned} y_{ik}^{MB_1}|P_{ik} &\sim \text{Bernoulli}(P_{ik}) \\ \text{logit}(P_{ik}) &= \beta_0 + f(\tilde{w}_{ik}) + u_k + v_k, \end{aligned} \tag{13}$$

or

$$\begin{aligned} y_{ik}^{MB_2}|P_{ik} &\sim \text{Bernoulli}(P_{ik}) \\ \text{logit}(P_{ik}) &= \beta_0 + f(\tilde{\pi}_{ik}) + u_k + v_k, \end{aligned} \tag{14}$$

where $f(.)$ is a non-parametric function in either the design weights or the sample probabilities, specified by a random walk model of order one (RW1) or a penalised spline (SP).

As we assume that not all information on the design is made available for the researcher (only the design weights for the sampled individuals), Vandendijck *et al.* [13] proposed a method to resample weights for the non-sampled individuals. Based on the work of Si *et al.* [35], a Bayesian model was developed in order to estimate the design weights of non-sampled individuals. For this model, we divide the data into L_k strata, whereby L_k is the number of unique design weights in area k. Denote n_{lk} as the sample size in poststratification cell l ($l = 1,...,L_k$) in area k and N_{lk} as the corresponding population size. Under the assumption of independent sampling, one can model the sampling probabilities for a given individual i in area k with a Bernoulli distribution, whereby the probabilities are given by

$$P(R_{ik} = 1) = \frac{c_k}{w_{ik}}. \tag{15}$$

In this formula c_k fulfills the roll of a positive normalising constant, in order to ascertain that the probabilities sum up to sample size n_k. Given the fact that all individuals in poststratification cell l have the same weight, we can define $w_{ik} \equiv w_{(l)k}$ and the expected value of n_{lk} can be expressed as $E(n_{lk}) = c_k \frac{N_{lk}}{w_{(l)k}}$. Since $n_k = \sum_{l=1}^{L_k} n_{lk}$, we can represent the normalising constant as $c_k = n_k \frac{1}{\sum_{l=1}^{L_k} \frac{N_{lk}}{w_{(l)k}}}$.

The vector of sample sizes in area k $(n_{1k},...,n_{L_k k})$ is assumed to follow a multinomial distribution, conditioned on n_k for each k:

$$(n_{1k},...,n_{L_k k}) \sim \left(n_k; \frac{\frac{N_{1k}}{w_{(1)k}}}{\sum_{l=1}^{L_k} \frac{N_{1k}}{w_{(1)k}}}, ..., \frac{\frac{N_{lk}}{w_{(l)k}}}{\sum_{l=1}^{L_k} \frac{N_{lk}}{w_{(l)k}}} \right). \tag{16}$$

In the above parametrisation, the population sizes N_{lk} are unknown parameters. Because of the fact that they are also unnormalised, we normalise them in such a way that they sum to the population size in area k after fitting them:

$$\tilde{N}_{lk} = \frac{N_{lk}}{\sum_{l=1}^{L_k} N_{lk}} N_k. \tag{17}$$

After obtaining information on $\tilde{N}_{lk}$, (12) can be rewritten as

$$\hat{P}_k = \frac{1}{\sum_{l=1}^{L_k} \tilde{N}_{lk}} \left(\sum_{l=1}^{L_k} n_{lk} \bar{y}_l + \sum_{l=1}^{L_k} \left(\tilde{N}_{lk} - n_{lk} \right) \hat{P}_{lk} \right), \tag{18}$$

where we define $\bar{y}_l = \frac{\sum_{i \in l} y_{ik}}{n_{lk}}$. An estimate for the prevalence P_{lk} in each poststratification cell l and area k can be obtained from (13) using the unique normalised weights $\tilde{w}_{lk}$. While this approach yields a point estimate for the population prevalence P_k, inference is based on the posterior distribution of $\hat{P}_k$. This posterior distribution is constructed by taking samples from the posterior distribution of $\tilde{N}_{lk}$ and $\hat{P}_{lk}$. These samples can be inserted B times in (18) in order to get B posterior samples for $\hat{P}_k$.

2.4 Adjustments for incomplete data

When dealing with nonresponse, weights need to be adjusted to take into account the reduced sample size and the possible imbalance due to missingness. A first approach is to work with the complete data as they are and normalise the weights according to the observed sample size m_k in each area. This leads to the use of the following weights

$$\tilde{w}_{ik}^{d'} = m_k \cdot \frac{w_{ik}^d}{\sum_{i \in s_k^*} w_{ik}^d}. \tag{19}$$

Because we only adjust the sample size in this case and leave the design weights unchanged, this procedure will be called the semi-adjusted method. However, since missingness might lead to an imbalance of the observations, e.g. when missingness is related to any of the design variables, it is important to re-weight observations using poststratification. This can be done by defining a new weight w_{ik}^*, defined as the product of the design weights w_{ik}^d and the missingness weight w_{ik}^m. This new missingness weight w_{ik}^m can be defined as $\frac{1}{P(r_{ik}=1)}$, whereby we can model this probability using the following logistic regression model

$$\text{logit}(P(r_{ik}=1)) = \alpha + \beta X_{ih}, \tag{20}$$

where X_{ih} is a vector containing information on the h covariates which might have an effect on the missingness process for individual i in area k. We can characterise the missingness weights as $w_{ik}^m = \frac{1+\exp(\alpha+\beta X_{ih})}{\exp(\alpha+\beta X_{ih})}$. The final weight is then normalised in such a way that it corresponds to the number of non-missing observations:

$$\tilde{w}_{ik}^* = m_k \cdot \frac{w_{ik}^d \cdot w_{ik}^m}{\sum_{i \in s_k^*} w_{ik}^d \cdot w_{ik}^m}. \tag{21}$$

Since we adjust both the sample size as well as the design weights for missingness, this procedure will be called the adjusted method. All spatial smoothing methods as discussed in Section 2.1 and 2.2 can be adapted to the setting of incomplete data by substituting the weight used in the HT estimator as either the semi-adjusted or the adjusted weight. Note that this weighting approach assumes that missingness does not depend on the unobserved outcome itself, and thus that missingness is not MNAR.

The hierarchical weight-smoothing model allows for an additional correction to account for nonresponse. Since the normalised weights are used directly as a covariate effect, this model allows us to discern between the effects of the design weights and the missingness weights. This separation can be expressed as follows:

$$\begin{aligned} y_{ik}^{MB_3} | P_{ik} &\sim \text{Bernoulli}(P_{ik}) \\ \text{logit}(P_{ik}) &= \beta_0 + f_1(\tilde{w}_{ik}^d) + f_2(\tilde{w}_{ik}^m) + u_k + v_k. \end{aligned} \tag{22}$$

This model has the advantage to be able to separate the impact of the design variables with those variables that affect the missingness probability. Indeed, the variables that explain the nonresponse may not be the same as the design variables. We can expand the previous model even further by adding an overdispersion parameter ε_{ik}, with $\varepsilon_{ik} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. This parameter will account for all remaining variability for individual i of district k, further improving the prediction of the area-specific prevalence (11). Thus, model (22) could be extended to:

$$\begin{aligned} y_{ik}^{MB_3} | P_{ik} &\sim \text{Bernoulli}(P_{ik}) \\ \text{logit}(P_{ik}) &= \beta_0 + f_1(\tilde{w}_{ik}^d) + f_2(\tilde{w}_{ik}^m) + u_k + v_k + \varepsilon_{ik}. \end{aligned} \tag{23}$$

The behavior of these methods under different degrees of missingness will be investigated in the next section.

2.5 Prior specification

All area-level and unit-level methods described in Sections 2.2, 2.3 and 2.4 were implemented using the Integrated Nested Lagrange Approximation (INLA) approach, described by Rue *et al.* [36]. INLA was implemented as an R-package and can be downloaded at http://www.r-inla.org/. It serves as a faster alternative to Markov Chain Monte Carlo (MCMC) methods when performing statistical inference for latent Gaussian models. INLA computes accurate approximations to the posterior marginals via numerical integration. Carroll *et al.* [37] performed an in-depth comparison in the ability to recover estimates between INLA and OpenBUGS.

Prior distributions for the parameters $\beta_0, \sigma_u^2, \sigma_v^2$ and σ_w^2 need to be specified. In general vague priors are preferred in order to minimize their effect on the inferential evaluation, as was investigated by Browne and Draper [38] and Gelman [39]. We assume a zero-mean normal distribution with a high variance for the baseline parameter β_0. Furthermore, we assign a Gamma(0.5, 0.008) prior for both the spatial and non-spatial precision parameters σ_u^{-2} and σ_v^{-2}, similar to Mercer *et al.* (2014). Lastly, we consider the prior distribution for σ_w^{-2} to be Gamma(1,0.01), in accordance with Wakefield (2009).

Furthermore, as a sensitivity analysis, we investigated the effect of the neighborhood structure. As mentioned in Section 2.2, a first-order neighborhood structure was used in the analysis. In order to investigate the robustness of the estimates with respect to the definition of the neighborhood structure, an additional neighborhood scheme was considered. Hereby, we consider two areas i_1 and i_2 to be neighbors if they share a common boundary or if they both share a boundary with a common neighbor i_3.

3 Simulation Study

The performance of all models described above are investigated via a simulation study. Section 3.1 describes the design of the simulation study in which we investigate the performance of the models under different missing data mechanisms. Section 3.2 summarises the results of the simulations.

3.1 Design of simulation study

The 43 administrative districts of Belgium, with a total population size of around ten million, were chosen to be the geographical setting of interest (see Figure 1). The population data was stratified into 18 age-groups, each defined by a five-year interval. The total population size and average age in the population is presented in Figure 1 (upper panels). The indicator x denotes the 18 different age-groups in which the Belgian population is categorised ($x = 1$ for ages $0-4$, $x = 2$ for ages $5-9$, ..., $x = 18$ for ages 85+), and $Y_{i(j)k}$ is the binary response variable for the i^{th} individual belonging to age stratum j in district k ($i = 1,...,N_k, j = 1,...,18, k = 1,...,43$).

Simulating Population Prevalences

We assume that the binary response variable follows a bernoulli distribution

$$Y_{i(j)k} \sim \text{Bernouilli}(P_{jk}),$$

where P_{jk} is the population prevalence in stratum j of district k. The following two models are considered for the simulation of the population prevalences:

$$(M1): \text{ logit}(P_{jk}) = \text{logit}(0.10) + 0.30 \cdot x_{i(j)k},$$
$$(M2): \text{ logit}(P_{jk}) = \text{logit}(0.10) + 0.30 \cdot x_{i(j)k} + u_k + v_k,$$

with $x_{i(j)k}$ the age category of individual i in area k. Compared to $(M1)$, a convolution term $u_k \sim \text{ICAR}(0, \sigma_u^2)$ with precision $\sigma_u^{-2} \sim \text{Gamma}(1.0, 0.5)$ was added in $(M2)$, encompassing an uncorrelated random effect $v_k \sim \mathcal{N}(0, 0.10)$. The spatial random effects were generated using INLA. The values for these random effects were held constant across all simulations, allowing us to investigate the prediction of the underlying spatial trend. The true district-specific prevalence can then be calculated by averaging the simulated prevalences P_{jk}, weighted by their corresponding population sizes N_{jk}:

$$P_k = \frac{\sum_{j=1}^{J} N_{jk} P_{jk}}{\sum_{j=1}^{J} N_{jk}}.$$

These are presented in Figure 1 (lower panels). It can be observed that, in both prevalence models, there is some degree of spatial heterogeneity.

Simulating Survey Sample

A survey sample of size 5000 is taken from the simulated population using a stratified probability design, according to the following procedure:

1. The sample size per area is taken proportional to the population size in each area (N_k). A multinomial distribution is employed in order to generate the sample sizes n_k per district in order to ensure the aforementioned proportionality:

$$(n_1, \ldots, n_K) \sim \text{Multinomial}\left(5000; \frac{N_1}{\sum_{k=1}^{K} N_1}, \ldots, \frac{N_k}{\sum_{k=1}^{K} N_k}\right).$$

Note that the sampling procedure depends solely on the population sizes of the districts, not on the spatial distribution of the simulated outcome.

2. Next we distribute these samples across the different strata within a district k. We denote q_{jk} as the selection probability stratum j is selected in district k. In this setting we assume that this probability depends on the age of the individuals, assuming older individuals have a higher probability of being sampled. Defining x_{jk} as the age group in stratum j and district k, this could be expressed as follows:

$$q_{jk} = \frac{\log(x_{jk}+1)}{\sum_{j=1}^{J} \log(x_{jk}+1)}.$$

The stratum-specific sample size n_{jk} within each district are consequently simulated using a multinomial distribution:

$$(n_{1k}, \ldots, n_{Jk}) \sim \text{Multinomial}(n_k; q_{1k}, \ldots, q_{Jk}).$$

3. Finally, we generate the number of cases within this sample using a binomial distribution according to the presumed population prevalences:

$$y_{jk} \sim \text{Bin}(n_{jk}, P_{jk}).$$

Note that, for a simulated sample, the survey design weight of observation i in district k, $w_{ik}^{d} \equiv w_{i(j)k}^{d}$, is calculated as the inverse of the proportion of observations that are sampled from the population in stratum j within district k.

Simulating Nonresponse

As we want to investigate the effect of nonresponse on the estimates' spatial trend, different missing data mechanisms will be considered. It is assumed that only $(1-\beta)\%$ of the respondents answer the question of interest, with $\beta = (0, 20, 40, 60)\%$. Different missing data mechanisms can be underlying this, and the assumed scenarios are summarised by the probability weights $q_{i(j)k}^{m}$ in Table 1. The probability to have not observed the response for the i^{th} individual in area k is then equal to $P(r_{i(j)k} = 0) = \beta \frac{q_{i(j)k}^{m}}{\sum_{j=1}^{18} q_{i(j)k}^{m}}$.

In (S1), no nonresponse is present in the data; all outcomes are observed. These results will be compared with the different scenarios in which some of the outcomes are unobserved. In (S2), some of the outcomes are missing completely at random (MCAR). This means that the observed outcomes are a random sample from the set of individuals that are contained in the survey. The amount of missingness is given by a fixed parameter β. In (S3)-(S6), the missing data mechanism is Missing at Random (MAR), in the sense that the probability of having a missing response depends on age. Here, it is assumed that missingness increases (S3) or decreases (S5) with age. In the appendix, some other age-related missingness mechanisms were considered and their results were displayed. In scenario (S4) and (S6), we additionally assume that the amount of missingness is spatially varying (S-MAR), incorporating a spatial random effect u_k in the missingness probability which follows a zero-mean ICAR model. Finally, in (S7) and (S8), we consider the setting where the missing data mechanism is missing not at random (MNAR), assuming that the amount of missingness depends on the outcome of interest. In (S8) a spatial random effect was added, adding extra variability into the sampling scheme. For the simulation of the spatial random effect in (S4), (S6) and (S8), a Gamma$(0.5, 0.008)$ distribution was used for the spatial precision parameter σ_u^{-2}, similar as in Mercer *et al*. [12] and Chen *et al*. [11]. The simulated spatial random effect is presented in Figure 2 in the appendix.

The weights which adjust for missingness, $w_{ik}^m \equiv w_{i(j)k}^m$, are computed by the reciprocal of the sample size within each stratum and district of the dataset wherein the observations with a missing response are excluded and the sample size in the matching stratum and district in the original data set with no missing values. The final weights which are used in the analysis, taking into account both the sampling design and nonresponse, are defined as $w_{ik}^* = w_{ik}^d \cdot w_{ik}^m$. This implies that every observation i in stratum j and district k has the same weight.

3.2 Simulation Results

For each combination of a prevalence model and missing data scenario we run $S = 100$ simulations. The results of the unweighted estimator (2), Horvitz-Thompson estimator (3), unadjusted binomial model (5), logit normal model (7), arcsin root normal model (8), pseudo-likelihood model (9), effective sample size method (10) and the hierarchical weight-smoothing models (12), (13), (16) and (17) are discussed in this section. For each of these expressions both the semi-adjusted (14) and adjusted (15) weights are considered.

In the presence of missing data, the weights need to be redefined. As explained in Section 2.3, we can either use the semi-adjusted weights $\tilde{w}_{ik}^{d'}$, which correct for the number of respondents, or use the adjusted weights $\tilde{w}_{ik}^*$ defined as function of design weight and missingness weight. Figure 2 shows the effect of different definitions of the weight on the bias of the area-specific prevalence estimates. The box plot corresponds with the bias of the area-specific prevalences for the 100 simulations. These results clearly indicate that the definition of the weights can have a serious impact on the results. While results between the semi-adjusted and adjusted weights are similar under the MCAR missing data mechanism, there is a large discrepancy between these under the scenario of MAR. This is not unexpected since the age-distribution of the sample is distorted when missingness is MAR, but not when missingness is MCAR. This indeed indicates that post-stratification of the weights for important covariates is very important when missingness occurs in the sample.

To evaluate the estimates, the squared bias and mean squared error (MSE) are used. $\hat{P_k}^{(s)}$ represents the estimated prevalence for area k, based on the s^{th} simulated sample, and $\bar{\hat{P_k}} = \frac{1}{S}\sum_s \hat{P_k}^{(s)}$ is the corresponding averaged value over all simulated surveys. The overall squared bias and MSE are then defined as:

$$\begin{aligned}
\text{Bias}^2 &= \frac{1}{K}\sum_{k=1}^{K}\left(\bar{\hat{P_k}} - P_k\right)^2, \\
\text{Variance} &= \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{S-1}\sum_{s=1}^{S}\left(\hat{P}_k^{(s)} - \bar{\hat{P_k}}\right)^2\right), \\
\text{MSE} &= \text{Bias}^2 + \text{Variance}.
\end{aligned}$$

Furthermore, we calculate the nominal coverage probabilities of the estimated prevalences. Tables 2, 3 and 4 summarise the results under the simulated model M1 (model with age-trend only), whereas Tables 5, 6 and 7 correspond to the prevalence scenario M2 (model with additional spatial heterogeneity, not explained by the covariate). These tables only include results for adjusted weights.

The results show that, as expected, the UNW estimator can have very large bias. The larger the discrepancy between sample and population, the larger the bias. Note that in some scenarios the missingness slightly corrects for the imbalance between sample and population, decreasing the bias in the sample. In general, however, this estimator cannot be recommended. Also the MSE of this estimator is very large, in general, as a result of large variability. As this estimator is a direct estimator, making use only of the information within an area, this estimator can be very unstable, especially for areas with a small sample size.

Overall, the HT estimator, making use of the adjusted weights, performs much better in terms of bias. Not only in the situation of complete data, but also when data are incomplete, bias is small. In the setting of missingness not at random however (S7 and S8), the bias of the HT estimator is increased again. This is not unexpected, as the weighting approach makes the assumption of missingness at random (indeed, missingness probabilities are assumed to be independent of the missing outcomes themselves). Looking at the MSE, it can be observed that MSE is often large for the HT estimator. This is again due to the fact that the estimator can be unstable when sample size is small, as only information within an area is being used.

Comparing the different indirect modeling approaches (logit normal model, arcsine root normal model, pseudo-likelihood model, effective sample size method and weight-smoothing model) versus the direct estimators, it can be observed the indirect approaches outperform the direct estimators, both in average squared bias, average MSE and the coverage probabilities of the 95% credible interval. The average squared bias of all indirect estimators is small, as long as missingness is MCAR or MAR. When the underlying prevalence process is spatially structured, the MB and PL estimators outperform all other methods, leading to smallest bias, MSE and best coverage for all scenarios. However, when the underlying prevalence process is spatially unstructured, all area-level and unit-level methods behave similar in terms of bias and MSE. Overall, MAR and MNAR leads to a deflation of the coverage for the AS, PL and ES models. Coverage for the area-level and unit-level methods is good for MAR scenarios. Further note that even in the situation that no spatial heterogeneity term is present in the true prevalence model, both the area-level and unit-level models including the spatial heterogeneity term improve the fit in terms of both squared bias and MSE. This shows that, indeed, the shrinkage of extreme local estimates, by the use of the area- and unit-level estimators, is advantageous in the small area situation.

Different versions of the MB method were considered, either using a spline or random walk (RW1) in the prediction model. Only small differences are observed between MB1 based on spline (SP) or RW1, though, in general, the squared bias based on the spline model is slightly larger than based on the RW1 model. However, the MSE is smallest for the spline model as it leads to smallest variability. When the prevalence model does not contain any spatial random effect (model M1), we do not see any improvement of model MB3 (using separate weights with or without overdispersion) as compared to MB1. However, when the prevalence model does contain a spatial random effect (model M2), a small improvement is observed. This can be due to the increased variability in the prevalence model, which can be better modeled via the extended models.

When the missingness process is spatially structured, the bias in the area-specific prevalences increases. Note that the S-MAR reflects a situation in which the missingness probabilities depend on an (unobserved) environmental factor. In the current analyses, only observed covariates were taken into consideration in the missingness probabilities, and therefore are mis-specified. This explains why there is an increase in bias for all methods in the S-MAR scenario as compared to the MAR scenario. While the increase in bias can be large for the direct estimators, this increase is limited for the area-level and unit-level estimators, as they further model the spatial heterogeneity. Incorporation of a spatial heterogeneity term in the model for the missingness probabilities might further improve the estimation.

In Figures 3 and 4, we present the spatial trends for the HT, AN and MB1 (based on RW1) models, under prevalence model M1 and M2, respectively. Here, missingness scenarios S2 and S6 are of interest, whereby 20% nonresponse was simulated. When comparing the estimates in Figure 3 to the true proportion in Figure 1 (bottom left panel), it is apparent that the HT, AN and MB1 (RW1) methods have difficulties to recover the true proportion in the southern districts. In those areas, lower sample sizes were acquired, as these were generated proportionally to the population sizes (Figure 1 (top left panel)). This renders the design-based estimator HT less reliable and inefficient, while the AN and MB1 estimators retrieve the true population proportion significantly better in the MCAR setting. However, when looking at the results of the S2 and S6 mechanisms in Figure 4 for the spatial prevalence model, all three estimators perform well.

In addition, Figures 3-6 show the estimated trends for simulation scenarios S1-S4, under both prevalence models (M1 and M2). Here the weights adjust for 60% missingness. These results can be found in the appendix.

A sensitivity analysis of the priors of the random effects is presented in the appendix. Tables 8-11 show the summary statistics for the models which assign a Gamma(2, 1) prior for the precision parameters σ_u^{-2} and σ_v^{-2}. Similarly, Tables 12-15 display the results when considering the prior distribution of σ_u^{-2} and σ_v^{-2} to be a Gamma(1, 0.5). Furthermore, we display the summary results for the analyses using the second-order neighboring structure in Tables 16-19. Overall, while one can detect small deviations across the different simulation settings, the results show that the models perform consistently when applying small deviations for the prior distributions and neighborhood structure.

4 Application to the Belgian HIS data set

We apply the methods described in Section 2 to the Belgian Health Interview Survey (HIS) (2001). The Belgian HIS aims to investigate the health status of the Belgian population. The selection process of the respondents comprises three steps. Firstly, respondents are selected based on the region in which they live (Flanders, Wallonia and the Brussels region). A total sample size of 12.770 was drawn from the population (4255, 3234 and 5281 respectively for the three regions). Secondly, a stratification is carried out at the level of the ten provinces. Lastly, within the provinces, the sampling units are selected in three stages. Municipalities are selected proportionally to their size within the provinces and form the primary selection unit (PSU). Within these municipalities a simple random sample of households was drawn, forming the secondary sampling unit (SSU). Lastly, not more four individuals, the tertiary sampling unit (TSU), were interviewed in each household.

For this analysis the variable "perceived health" was investigated. Participants answered the question "How is your health in general?" on a scale from "Excellent" to "Very Poor". This ordinal variable was combined into two groups, using the dichotomization which can be conferred in Table 20. People which answered to be feeling "Very Good" or "Good" were allocated value 0, while people who responded to be feeling "Reasonable" or worse got value 1 for the newly constructed variable. This allows us to use the methods, outlined in Section 2.

The data collection was carried out over the same 18 age groups that were used in Section 3. Only participants who were older than 15 were taken into account for the analysis as this question was only asked to individuals older than 15. The line plot in Figure 5 depicts a positive correlation between the age of the participants and the perceived health, as well as with the amount of missingness, after age 15.

In total, 10419 eligible observations were taken into account for the analysis. When using the constructed binary perceived health variable, 2383 participants (22.87%) stated that their general health was reasonable or worse, 6998 respondents (67.17%) found their health in general to be good or better. 1038 (9.96%) people did not answer this question. The sample sizes in the different Belgian districts range from 43 to 2576. The amount of missingness varies between 2% and 23 %, where the districts Ath and Mons in the province of Hainaut exhibit the largest degrees of incomplete data. These results are visualised in Figure 7. Four districts (Dendermonde, Dinant, Ieper, Veurne) were not sampled in the HIS study and are highlighted in black. The original design weights were poststratified in order to account for any distributional differences for age and gender. We can define the poststratified design weights $w^d_{i(j)k}$ as follows:

$$w^d_{i(j)k} = w_{i(j)k} \cdot \frac{N_{jk}}{\sum_{i\in(j)} w_{i(j)k}},$$

where $w_{i(j)k}$ and $w^d_{i(j)k}$ are the original design weight and the poststratified design weight respectively for individual i in stratum j and area k. The unnormalised missingness weights were constructed using the following logistic regression models:

$$(W1):\ \text{logit}(P(r_{ik}=1)) = \eta_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Gender}_i,$$
$$(W2):\ \text{logit}(P(r_{ik}=1)) = \eta_i = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{Gender}_i + u_k + v_k,$$

We model the probability of observing the response of interest for individual i in area k in terms of the age and gender (gender_i = 0 for males and gender_i = 1 for females) of the given respondent. As we also want to investigate whether methods can be further improved by incorporating a spatial trend in the weights, we extended the model by incorporating a spatial (u_k) and non-spatial (v_k) random effect in the estimation process which will account for any spatially correlated and uncorrelated variability not yet explained by the design variables Age and Gender. Finally, we can again characterise the missingness weights as $w^m_{ik} = \frac{1+\exp(\eta_i)}{\exp(\eta_i)}$.

The normalised design weights vary from 0.3865 to 5.595 while the normalised weights which adjust for nonresponse range between 0.0282 and 3.575.

4.1 Data Application

The right panel in Figure 5 shows the box plots for comparing the estimates associated with the adjusted and semi-adjusted weights obtained by the models described in Section 2. Furthermore, the hierarchical smoothing models, whereby the non-parametric function is specified in terms of the two normalized weights are visualized as well. One could observe that the effect of the adjustment of the weights is limited in the estimation process. This will be further investated by means of a simulation study in the next section.

We can also evaluate the unit-level methods based on their goodness-of-fit. In Table 23 the deviance information criterion (DIC) values for each of the models are shown. It could be observed that the approaches which use the adjusted weights in the estimation process perform slightly better, compared to those which make use of the semi-adjusted weights. Furthermore, we can observe that incorporation of the spatial heterogeneity into the missingness weights (W2) lead to higher DIC values as compared to the methods using weights that do not incorporate the spatial heterogeneity (W1). This is an indication that the spatial trend of the nonresponse does not play a major role in this application. Because the semi-adjusted weights do not adjust for the missingness in the data, no estimates for the MB3 models were obtained, since the missingness weights can not be estimated. Also, we could not compute the DIC statistics for the MB3 (SP+OD) and MB3 (SP) methods under the W1 and W2 approach respectively. For these two situations, INLA could not allocate enough memory in order to perform the estimations. It should be noted that, with the exception of the five MB estimators, the DIC values cannot be compared among the different area-level methods as the outcome of interest varies between them. When looking at the DIC values of the hierarchical weight-smoothing models, we can conclude that the MB3 models perform best in terms of goodness-of-fit.

Figure 8 presents the spatial distribution of the prevalence for six modeling approaches: The unweighted estimator, the Horvitz-Thompson estimator, the arcsine root normal estimator, the pseudo-likelihood estimator and two hierarchical weight-smoothed methods: one whereby the weights are modeled through a non-parametric function which is specified by a random walk, while the other utilises a penalised spline and an overdispersion parameter. The spatial trend based on the UNW and HT estimator are most variable. Note that no estimates could be obtained for the areas in which no samples were taken. As such, these areas are colored in black. In addition, it can be observed that those areas where the sample size is smallest or missingness is highest have the most extreme prevalence estimates. This instability was also observed in the simulation study. The estimated spatial trend based on the indirect estimators behave in a similar way. There seems to be an important North-South trend, with higher prevalence of poor perceived health in the Southern part of Belgium. In this analysis, all methods lead to very similar results. In order to understand the robustness of the methods with respect to the design and missingness, we performed an additional simulation study.

4.2 Simulation Setting

In the simulation study, the stratified clustered multi-stage design of the HIS data set was kept intact. The survey sample has the same amount of observations as the 2001 Health Interview Survey, namely 10419. The sample size and its distribution across the different strata remain the same over all simulation runs. For each simulation run, we assume that the response variable is missing. Consequently we use the following approach to simulate the response.

Simulating Prevalences

We assume that the binary response variable follows a bernoulli distribution

$$Y_{ik} \sim \text{Bernouilli}(P_{ik}),$$

where P_{ik} denotes the population prevalence for individual i of district k. The following two models are considered for the simulation of the population prevalences:

$$\text{logit}(P_{ik}) = \eta_{ik} = \hat{\beta}_0 + \hat{\beta}_1 * \text{Age}_{ik} + \hat{\beta}_2 * \text{Gender}_{ik} + u_k + v_k$$

A convolution term $u_k \sim \text{ICAR}(0, \sigma_u^2)$ with precision $\sigma_u^{-2} \sim \text{Gamma}(1.0, 0.5)$ was used, encompassing an uncorrelated random effect $v_k \sim \mathcal{N}(0, 0.10)$. The spatial random effects were generated using INLA. The values for these random effects were held constant across all simulations. We also include an additional individual-specific random effect ε_{ik} which follows a $\mathcal{N}(0, 0.10)$-distribution as well. The parameter estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ were yielded from the same logistic model, performed on the original data set. We assume a bernoulli distribution when simulating the outcome variable Y_{ik}^*:

$$Y_{ik}^* = \text{Bern}(P_{ik})$$

Simulating Nonresponse

The probability of not observing the response for the i^{th} individual in area k is then equal to $P(r_{ik} = 0) = \beta \frac{q_{ik}^m}{\sum_{i,k} q_{ik}^m}$, whereby $\beta = 0.10$ (conform to the overall nonresponse percentage in the HIS data set) and Table 21 contains the values of q_{ik}^m. By using these values β and q_{ik}^m, we aim to simulate the missingness pattern across the age groups as seen in Figure 7 (right panel).

Results

In the analysis of the simulated data sets, we will focus primarily on the (W2) scenario when estimating the missingness weights as this approach will account for more variability than (W1). Figure 6 displays the area-specific estimates of the 100 simulation runs whereby the adjusted, semi-adjusted and separate weights were applied in the estimation process. It is apparent that the effect of adjusting for nonresponse is fairly limited. An explanation could be found in the fact that the missingness distribution does not express a distinct trend across the eligible age groups and that there is a low amount of missingness overall. When performing the analysis, INLA indicated having difficulties allocating memory for the analyses of the MB3 models using a penalised spline, both with or without the overdispersion term. This caused some of simulation runs to abort the estimation process prematurely for these methods. However, the data sets for which the models did converge, showed a clear bias as can be seen in Figure 6. As such, when performing an analysis with these latter two models, one has to be cautious when interpreting the results. The summary statistics for the adjusted weights (W2) are provided in Table 22. Overall, the HT and AN perform best in terms of squared bias, closely followed by the MB1 models. When looking at the MSE statistics, the unit-levels MB1 and MB3 (RW1) models perform considerably better than the area-level models. And thus, it follows the same conclusion as the previous simulation study in Section 4. As mentioned early, the unit-level models MB3 models whereby a penalised spline was applied performs considerably worse than the other approaches.

Note that the mechanism of simulation used in this section is different from the earlier simulation study, as the current simulation study does not involve any sampling variability. This exercise however still shows that models should properly reflect features of the sampling design, otherwise inferences are likely to be distorted.

5 Conclusion and Discussion

In this paper, the impact of missingness in health surveys on the estimation of the area-specific prevalences was investigated. A comparison of different methods to estimate the area-specific prevalences were compared, ranging from an unweighted estimator, the Horvitz-Thompson estimator, to area-level and unit-level approaches taking into account spatially structured random effects. Weighting methods were preferred in this paper, as weighting can be used to account for both the design of the survey as well as for missingness. Weighting methods are only valid under the assumption of missing completely at random and missing at random, confirmed by the simulation that pointed to increased bias in the case of missingness not at random. The area-specific prevalences are well estimated based on the weight smoothing methods, taking into account the design and missingness weights as a covariate in the model and accounting for possible spatial correlation via a convolution model. Also coverage is very well retained for the unit-level weight smoothing methods. If missingness is spatially structured, this has a negative impact on the prevalence estimation, leading to slightly increased bias and MSE. In conclusion, the use of weight-smoothing methods accounting for poststratification weights to account for incompleteness in the data are very promising when estimating the spatial trend based on survey data.

While common interest in this paper was on the estimation of the area-specific prevalence, the unit-level models also allow to study the risk after accounting for know risk factors. It would be of interest to study how a standardized rate can be obtained in the context of small area estimation. This is a topic of further research.

Another extension of the proposed method is the use of an alternative spatial prior. While the ICAR prior is commonly used in spatial modeling of lattice data, other spatial priors were proposed in literature. An interesting option is the use of the Leroux prior [40], as the Leroux prior has some advantages over the ICAR prior in terms of the ability to estimate the correlation effect.

Inference for all considered models (both area- and unit-level models) was done in the Bayesian framework. The traditional approach towards Bayesian inference is the use of MCMC (Markov Chain Monte Carlo methods). We however investigated INLA (Integrated Nested Laplace Approximations), as it serves as a faster alternative to MCMC methods when performing statistical inference for latent Gaussian models. Carroll *et al*. [37] performed an in-depth comparison in the ability to recover estimates between INLA and OpenBUGS in the context of spatial hierarchical modeling. Chen *et al*. [11] also showed a comparison between INLA and OpenBUGS for estimation of the ES model. It however remains to be investigated how INLA compares with more general MCMC methods in the specific context of missing data in small area estimation.

Acknowledgements

Support from the National Institutes of Health is acknowledged [award number R01CA172805]. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged. For the simulation study we used the infrastructure of the VSC - Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government - department EWI.

References

1. Elliott P, Wakefield JC, Best NG, Briggs DJ. *Spatial Epidemiology: Methods and applications.* London: Oxford University, 2001.

2. Waller IA, Gotway CA. *Applied spatial statistics for public health data.* Hoboken: John Wiley & Sons Inc, 2004.

3. Lawson AB. *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology, second edition.* Boca Raton: Chapman & Hall/CRC, 2013.

4. Schaeffer RL, Mendenhall W, Ott L. *Elementary survey sampling.* Boston: PWS-Kent Publishing Company, 1990.

5. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 1952; **47**(260): 663-685.

6. Basu D. Godambe VP, Sprott DA (Eds). *An essay on the logical foundations of survey sampling, Part One. In Foundations of Statistical Inference.* Toronto: Holt, Rinehart and Winsten, 1971; 202-233.

7. Rao JNK. Impact of frequentist and Bayesian methods on survey sampling practice: A selective appraisal. *Statistical Science* 2011; **36**(2): 240-256.

8. Gelman A. Struggles with survey weighting and regression modeling. *Statistical Science* 2007; **22**(2): 153-164.

9. Little RJA. Comment: Struggles with survey weighting and regression modeling. *Statistical Science* 2007; **22**(2): 171-174.

10. Pfeffermann D. Comment: Struggles with survey weighting and regression modeling. *Statistical Science* 2007; **22**(2): 179-183.

11. Chen C, Wakefield JC, Lumley T. The use of sampling weights in Bayesian hierarchical models for small area estimation. *Spatial end Spatio-Temporal Epidemiology* 2015; **11**: 33-43.

12. Mercer L, Wakefield JC, Chen C, Lumley T. A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spatial Statistics* 2014; **8**: 69-85.

13. Vandendijck Y, Faes C, Kirby RS, Lawson A, Hens N. Model-based inference for small area estimation with sampling weights. *Spatial Statistics* 2016; **18**: 455-473.

14. Rubin DB. *Multiple imputation for nonresponse in surveys.* New York: John Wiley & Sons Inc, 1987.

15. Little RJA, Rubin DB. *Statistical analysis with missing data.* New York: John Wiley & Sons Inc, 1987.

16. Rao JNK. *Small Area Estimation.* Hoboken: John Wiley & Sons Inc, 2003.

17. Pfeffermann, D. New important developments in small area estimation. *Spatial Statistics* 2013; **28**: 40-68.

18. Ugarte MD, Goicoa T, Militino AF, Durbn M. Spline smoothing in small area trend estimation and forecasting, *Computational Statistics and Data Analysis* 2009, **53**(10), 3616 3629.

19. Salvati N, Chandra H, Ranalli MG, Chambers R. Small area estimation using a nonparametric model-based direct estimator, *Computational Statistics and Data Analysis* 2010, **54**(9), 2159-2171.

20. Chambers R, Rao JNK, Morales D, Ugarte MD. Special issue on small area estimation, *Computational Statistics & Data Analysis* 2012, 2837-2839.

21. Rao JNK, Molina I. *Small area estimation*, 2nd edition, Wiley, 2015.

22. Fay RE, Herriot RA. Estimates of income for small places: an application of JamesStein procedures to census data. *Journal of the American Statistical Association* 1979; **74**(366): 269-277.

23. Besag J, York J, Molli A. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991; **43**(1): 1-59.

24. Rue H, Held L. *Gaussian Markov Random Fields: Theory and Applications.* Boca Raton: Chapman and Hall/CRC Press, 2005.

25. Lawson A. *Statistical Methods in Spatial Epidemiology*, 2nd Edition, Wiley, 2006.

26. Bivand RS, Pebesma EJ, Gómez-Rubio V. *Applied Spatial Data Analysis with R*. Springer, 2008.

27. Raghunathan T, Xie D, Schenker N, Parsons V, Davin W, Dood K, Feuer E. Combining information from two surveys to estimate county-level prevalance rates of cancer risk factors and screening. *Journal of the American Statistical Association* 2007; **102**(478): 474-486.

28. Efron B, Morris C. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* 1979; **70**(350): 311-319.

29. Congdon P, Lloyd P. Estimating small area diabetes prevalence in the US using the behavioral risk factor surveillance system. *Journal of Data Science* 2010; **8**: 235-252.

30. Asparouhov T. General multi-level modeling with sample weights. *Communications in Statistics - Theory and Methods* 2006; **35**(3): 439-460.

31. Royall R. On finite population sampling theory under certain linear regression models. *Biometrika* 1970; **52**(2): 377-378.

32. Zheng H, Little RJA. Penalized spline model-based estimation of finite population total from probability-proportional-to-size samples. *Journal of Official Statistics* 2003; **19**: 99-117.

33. Zheng H, Little RJA. Inference for the population total from probability-proportional-to-size samples based on predictions from a penalized spline nonparametric model. *Journal of Official Statistics* 2005; **21**: 1-20.

34. Chen XQ, Elliot MR, Little R. Bayesian penalized spline model-based inference for finite population poroportion in unequal probability sampling. *Survey Methodology* 2010; **36**(1): 23-34.

35. Si Y, Pillai NS, Gelman A. Bayesian nonparametric weighted sampling inference. *Bayesian Analysis* 2015; **10**(3): 605-625.

36. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society* 2009; **71**(2): 319-392.

37. Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spatial and Spatio-temporal Epidemiology* 2015; **14-15**: 45-54.

38. Browne JB, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* 2006; **1**(3): 473-514.

39. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006; **1**(3): 515-533.

40. Leroux BG, Lei X, Breslow N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. *Statistical Models in Epidemiology, the Environment and Clinical Trials*, Halloran M, Berry D (eds). Springer-Verlag: New York; 179-191, 1999.

Tables

Scenario	Description	Missingness probability weight	Overall probability
S1	No missing data	$q^m_{i(j)k} = 1$	$\beta = 0$
S2	MCAR	$q^m_{i(j)k} = 1$	$\beta = (0.2, 0.4, 0.6)$
S3	MAR	$q^m_{i(j)k} = 1 - \frac{x_{i(j)k}}{20}$	$\beta = (0.2, 0.4, 0.6)$
S4	S-MAR	$q^m_{i(j)k} = \text{expit}\left(\text{logit}\left(1 - \frac{x_{a,ik}}{20}\right) + u_k\right)$	$\beta = (0.2, 0.4, 0.6)$
S5	MAR	$q^m_{i(j)k} = \frac{x_{i(j)k}}{20}$	$\beta = (0.2, 0.4, 0.6)$
S6	S-MAR	$q^m_{i(j)k} = \text{expit}\left(\text{logit}\left(\frac{x_{a,ik}}{20}\right) + u_k\right)$	$\beta = (0.2, 0.4, 0.6)$
S7	MNAR	$q^m_{i(j)k} = 0.7^{y_{i(j)k}} 0.9^{(1-y_{i(j)k})}$	$\beta = (0.2, 0.4, 0.6)$
S8	MNAR	$q^m_{i(j)k} = \text{expit}\left(\text{logit}\left(0.7^{y_{i(j)k}} 0.9^{(1-y_{i(j)k})}\right) + u_k\right)$	$\beta = (0.2, 0.4, 0.6)$

Table 1: *Description of the simulated nonresponse mechanisms*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	16.86	0.65	0.45	0.27	0.33	**0.21**	0.26	/	/	/
						20% Missingness				
MCAR (S2)	16.86$^{(10)}$	0.87$^{(9)}$	0.62$^{(8)}$	0.37$^{(2)}$	0.45$^{(6)}$	**0.37**$^{(1)}$	0.43$^{(5)}$	0.38$^{(3)}$	0.41$^{(4)}$	0.47$^{(7)}$
MAR (S3)	10.96$^{(10)}$	0.48$^{(9)}$	0.36$^{(8)}$	0.24$^{(3)}$	0.29$^{(6)}$	**0.21**$^{(1)}$	0.27$^{(5)}$	0.23$^{(2)}$	0.27$^{(4)}$	0.30$^{(7)}$
S-MAR (S4)	11.27$^{(10)}$	0.54$^{(9)}$	0.40$^{(8)}$	0.26$^{(3)}$	0.32$^{(6)}$	**0.23**$^{(1)}$	0.30$^{(5)}$	0.24$^{(2)}$	0.29$^{(4)}$	0.32$^{(7)}$
MAR (S5)	12.34$^{(10)}$	0.62$^{(9)}$	0.43$^{(8)}$	0.27$^{(5)}$	0.33$^{(7)}$	0.26$^{(4)}$	0.30$^{(6)}$	0.13$^{(2)}$	**0.12**$^{(1)}$	0.17$^{(3)}$
S-MAR (S6)	12.31$^{(10)}$	0.58$^{(9)}$	0.39$^{(8)}$	0.25$^{(3)}$	0.30$^{(6)}$	0.25$^{(2)}$	0.28$^{(5)}$	**0.22**$^{(1)}$	0.28$^{(4)}$	0.31$^{(7)}$
MNAR (S7)	7.81	0.60	0.50	0.61	0.54	**0.30**	0.59	0.53	0.55	0.55
S-MNAR (S8)	8.00	0.58	0.47	0.57	0.49	**0.27**	0.55	0.51	0.52	0.52
						40% Missingness				
MCAR (S2)	16.63$^{(10)}$	1.14$^{(9)}$	0.83$^{(8)}$	**0.49**$^{(1)}$	0.61$^{(3)}$	0.57$^{(2)}$	0.66$^{(6)}$	0.63$^{(4)}$	0.64$^{(5)}$	0.69$^{(7)}$
MAR (S3)	4.77$^{(10)}$	0.22$^{(8)}$	0.21$^{(6)}$	**0.17**$^{(1)}$	0.22$^{(7)}$	0.21$^{(4)}$	0.26$^{(9)}$	0.19$^{(2)}$	0.21$^{(5)}$	0.20$^{(3)}$
S-MAR (S4)	5.11$^{(10)}$	0.24$^{(6)}$	0.24$^{(5)}$	0.20$^{(2)}$	0.22$^{(3)}$	0.25$^{(7)}$	0.32$^{(9)}$	**0.19**$^{(1)}$	0.25$^{(8)}$	0.23$^{(4)}$
MAR (S5)	7.03$^{(10)}$	0.38$^{(9)}$	0.30$^{(7)}$	**0.22**$^{(1)}$	0.27$^{(4)}$	0.27$^{(3)}$	0.32$^{(8)}$	0.26$^{(2)}$	0.29$^{(6)}$	0.28$^{(5)}$
S-MAR (S6)	7.02$^{(10)}$	0.34$^{(9)}$	0.29$^{(7)}$	**0.21**$^{(1)}$	0.24$^{(3)}$	0.23$^{(2)}$	0.30$^{(8)}$	0.27$^{(6)}$	0.25$^{(4)}$	0.25$^{(5)}$
MNAR (S7)	**0.88**	4.26	4.28	4.77	4.45	4.53	4.56	4.26	4.38	4.33
S-MNAR (S8)	**1.44**	4.14	3.89	4.36	4.00	4.14	4.17	3.47	3.97	3.97
						60% Missingness				
MCAR (S2)	16.36$^{(10)}$	1.72$^{(9)}$	1.28$^{(8)}$	**0.83**$^{(1)}$	0.98$^{(2)}$	1.12$^{(3)}$	1.25$^{(7)}$	1.17$^{(5)}$	1.14$^{(4)}$	1.21$^{(6)}$
MAR (S3)	0.38$^{(10)}$	0.18$^{(8)}$	0.09$^{(3)}$	0.11$^{(5)}$	0.14$^{(6)}$	0.18$^{(7)}$	0.20$^{(9)}$	0.10$^{(4)}$	**0.08**$^{(1)}$	0.08$^{(2)}$
S-MAR (S4)	0.60$^{(10)}$	0.15$^{(7)}$	0.10$^{(3)}$	0.12$^{(4)}$	0.13$^{(6)}$	0.20$^{(8)}$	0.23$^{(9)}$	0.13$^{(5)}$	0.10$^{(2)}$	**0.09**$^{(1)}$
MAR (S5)	1.99$^{(10)}$	0.10$^{(1)}$	0.11$^{(3)}$	0.13$^{(5)}$	0.15$^{(7)}$	0.20$^{(8)}$	0.23$^{(9)}$	0.15$^{(6)}$	0.12$^{(4)}$	**0.11**$^{(2)}$
S-MAR (S6)	2.13$^{(10)}$	0.13$^{(2)}$	0.13$^{(3)}$	0.13$^{(4)}$	0.15$^{(6)}$	0.19$^{(8)}$	0.23$^{(9)}$	0.17$^{(7)}$	0.14$^{(5)}$	**0.12**$^{(1)}$
MNAR (S7)	**5.60**	19.68	19.08	19.89	18.81	18.96	18.94	18.44	19.29	19.15
S-MNAR (S8)	**6.23**	19.00	17.45	18.09	17.06	17.22	17.23	16.52	17.57	17.55

Table 2: *Summary statistics of squared bias using adjusted weigths, analysed under the M1 simulation mechanism. The prior distribution of the precision parameters* σ_u^{-2} *and* σ_v^{-2} *are assumed to follow a Gamma(0.5,0.008)-distribution.* ($\times 10^3$)

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	20.58	6.22	1.61	**0.62**	0.69	0.83	0.76	/	/	/
	20% Missingness									
MCAR (S2)	21.65$^{(10)}$	7.99$^{(9)}$	2.01$^{(8)}$	**0.83**$^{(1)}$	0.92$^{(2)}$	1.03$^{(5)}$	0.95$^{(3)}$	1.04$^{(6)}$	0.95$^{(4)}$	1.38$^{(7)}$
MAR (S3)	15.80$^{(10)}$	7.04$^{(9)}$	1.47$^{(8)}$	**0.56**$^{(1)}$	0.60$^{(2)}$	1.03$^{(5)}$	0.74$^{(3)}$	1.18$^{(6)}$	1.03$^{(4)}$	1.23$^{(7)}$
S-MAR (S4)	16.07$^{(10)}$	6.85$^{(9)}$	1.51$^{(8)}$	**0.58**$^{(1)}$	0.63$^{(2)}$	0.94$^{(4)}$	0.75$^{(3)}$	1.18$^{(6)}$	1.04$^{(5)}$	1.25$^{(7)}$
MAR (S5)	17.03$^{(10)}$	7.21$^{(9)}$	1.58$^{(8)}$	**0.60**$^{(1)}$	0.65$^{(2)}$	0.96$^{(7)}$	0.76$^{(5)}$	0.87$^{(6)}$	0.73$^{(3)}$	0.75$^{(4)}$
S-MAR (S6)	17.10$^{(10)}$	7.18$^{(9)}$	1.59$^{(8)}$	**0.60**$^{(1)}$	0.64$^{(2)}$	1.00$^{(5)}$	0.75$^{(3)}$	1.06$^{(6)}$	0.95$^{(4)}$	1.26$^{(7)}$
MNAR (S7)	12.75	7.10	1.64	**0.95**	0.86	1.30	1.14	1.26	1.16	1.45
S-MNAR (S8)	12.95	7.25	1.71	**0.94**	0.92	1.24	1.12	1.27	1.11	1.43
	40% Missingness									
MCAR (S2)	22.94$^{(10)}$	10.41$^{(9)}$	2.50$^{(8)}$	**1.08**$^{(1)}$	1.25$^{(3)}$	1.26$^{(4)}$	1.23$^{(2)}$	1.31$^{(6)}$	1.28$^{(5)}$	1.70$^{(7)}$
MAR (S3)	11.75$^{(10)}$	9.00$^{(9)}$	1.39$^{(8)}$	**0.52**$^{(1)}$	0.52$^{(2)}$	1.20$^{(5)}$	0.69$^{(3)}$	1.28$^{(7)}$	1.12$^{(4)}$	1.25$^{(6)}$
S-MAR (S4)	11.62$^{(10)}$	8.26$^{(9)}$	1.41$^{(8)}$	**0.51**$^{(1)}$	0.59$^{(2)}$	1.14$^{(5)}$	0.74$^{(3)}$	1.32$^{(7)}$	1.12$^{(4)}$	1.24$^{(6)}$
MAR (S5)	13.92$^{(10)}$	9.11$^{(9)}$	1.51$^{(8)}$	**0.57**$^{(1)}$	0.58$^{(2)}$	1.05$^{(4)}$	0.79$^{(3)}$	1.19$^{(6)}$	1.06$^{(5)}$	1.31$^{(7)}$
S-MAR (S6)	13.80$^{(10)}$	8.63$^{(9)}$	1.56$^{(8)}$	**0.55**$^{(1)}$	0.64$^{(2)}$	0.97$^{(4)}$	0.74$^{(3)}$	1.23$^{(6)}$	1.03$^{(5)}$	1.31$^{(7)}$
MNAR (S7)	7.91	12.64	5.41	**5.12**	4.74	5.23	5.03	5.18	5.08	5.31
S-MNAR (S8)	8.52	12.69	5.18	**4.73**	4.39	4.85	4.69	4.42	4.68	4.96
	60% Missingness									
MCAR (S2)	26.09$^{(10)}$	15.47$^{(9)}$	3.13$^{(8)}$	**1.58**$^{(1)}$	1.60$^{(2)}$	1.84$^{(3)}$	1.87$^{(4)}$	2.40$^{(6)}$	2.33$^{(5)}$	2.74$^{(7)}$
MAR (S3)	11.44$^{(10)}$	13.01$^{(9)}$	1.56$^{(8)}$	**0.53**$^{(1)}$	0.62$^{(2)}$	1.04$^{(4)}$	0.71$^{(3)}$	1.40$^{(7)}$	1.13$^{(5)}$	1.25$^{(6)}$
S-MAR (S4)	11.92$^{(10)}$	12.91$^{(9)}$	1.34$^{(8)}$	0.52$^{(2)}$	**0.47**$^{(1)}$	1.07$^{(4)}$	0.74$^{(3)}$	1.43$^{(7)}$	1.14$^{(5)}$	1.29$^{(6)}$
MAR (S5)	12.98$^{(10)}$	12.85$^{(9)}$	1.41$^{(8)}$	0.54$^{(2)}$	**0.48**$^{(1)}$	0.99$^{(4)}$	0.72$^{(3)}$	1.27$^{(6)}$	1.08$^{(5)}$	1.34$^{(7)}$
S-MAR (S6)	13.05$^{(10)}$	12.66$^{(9)}$	1.60$^{(8)}$	**0.57**$^{(1)}$	0.65$^{(2)}$	0.96$^{(4)}$	0.73$^{(3)}$	1.30$^{(6)}$	1.11$^{(5)}$	1.38$^{(7)}$
MNAR (S7)	**16.72**	31.75	20.22	20.25	19.08	19.64	19.43	18.98	19.66	19.81
S-MNAR (S8)	18.11	31.63	18.69	18.59	17.44	18.07	17.93	**17.17**	18.05	18.29

Table 3: *Summary statistics of MSE using adjusted weigths, analysed under the M1 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(0.5,0.008)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	0.33	0.95	0.95	0.98	0.97	1.00	1.00	/	/	/
	20% Missingness									
MCAR (S2)	0.39	0.94	0.95	0.98	0.97	1.00	0.99	0.99	0.99	1.00
MAR (S3)	0.52	0.95	0.97	0.98	0.98	1.00	0.99	0.99	1.00	1.00
S-MAR (S4)	0.51	0.95	0.97	0.98	0.98	1.00	1.00	0.98	1.00	1.00
MAR (S5)	0.49	0.95	0.97	0.99	0.98	1.00	1.00	0.99	0.99	1.00
S-MAR (S6)	0.49	0.95	0.97	0.98	0.98	1.00	1.00	0.99	1.00	1.00
MNAR (S7)	0.63	0.92	0.92	0.87	0.89	0.96	0.95	0.98	0.98	0.99
S-MNAR (S8)	0.62	0.92	0.92	0.88	0.90	0.96	0.96	0.98	0.98	0.99
	40% Missingness									
MCAR (S2)	0.47	0.93	0.95	0.98	0.97	1.00	0.99	0.99	0.99	1.00
MAR (S3)	0.79	0.92	0.92	0.88	0.88	0.96	0.95	0.99	1.00	1.00
S-MAR (S4)	0.78	0.92	0.94	0.89	0.89	0.96	0.96	0.99	1.00	1.00
MAR (S5)	0.71	0.93	0.96	0.95	0.96	0.99	0.98	0.99	1.00	1.00
S-MAR (S6)	0.70	0.94	0.96	0.95	0.94	0.99	0.99	0.99	1.00	1.00
MNAR (S7)	0.92	0.75	0.61	0.29	0.29	0.58	0.50	0.74	0.72	0.85
S-MNAR (S8)	0.89	0.76	0.64	0.36	0.38	0.62	0.56	0.79	0.76	0.86
	60% Missingness									
MCAR (S2)	0.56	0.91	0.93	0.98	0.97	1.00	0.99	0.99	0.99	1.00
MAR (S3)	0.93	0.85	0.83	0.65	0.64	0.81	0.78	1.00	1.00	1.00
S-MAR (S4)	0.92	0.86	0.83	0.67	0.67	0.83	0.79	1.00	1.00	1.00
MAR (S5)	0.89	0.89	0.91	0.82	0.82	0.94	0.92	1.00	1.00	1.00
S-MAR (S6)	0.89	0.90	0.91	0.83	0.82	0.95	0.93	0.99	1.00	1.00
MNAR (S7)	0.81	0.53	0.17	0.02	0.01	0.11	0.07	0.29	0.30	0.46
S-MNAR (S8)	0.80	0.54	0.24	0.05	0.05	0.20	0.16	0.35	0.38	0.50

Table 4: *Nominal coverage probabilities using adjusted weigths, analysed under the M1 simulation mechanism. The prior distribution of the precision parameters* σ_u^{-2} *and* σ_v^{-2} *are assumed to follow a Gamma(0.5,0.008)-distribution.*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	**0.04**	0.04	0.89	0.33	1.76	0.33	0.33	/	/	/
					20% Missingness					
MCAR (S2)	**0.03**$^{(1)}$	0.04$^{(2)}$	1.68$^{(9)}$	0.51$^{(8)}$	2.52$^{(10)}$	0.49$^{(5)}$	0.50$^{(7)}$	0.49$^{(4)}$	0.50$^{(6)}$	0.48$^{(3)}$
MAR (S3)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.37$^{(9)}$	0.46$^{(4)}$	2.01$^{(10)}$	0.47$^{(6)}$	0.47$^{(7)}$	0.54$^{(8)}$	0.47$^{(5)}$	0.45$^{(3)}$
S-MAR (S4)	**0.04**$^{(1)}$	0.04$^{(2)}$	1.49$^{(9)}$	0.50$^{(5)}$	2.40$^{(10)}$	0.50$^{(7)}$	0.50$^{(8)}$	0.47$^{(3)}$	0.50$^{(6)}$	0.49$^{(4)}$
MAR (S5)	**0.02**$^{(1)}$	0.03$^{(2)}$	1.60$^{(9)}$	0.44$^{(3)}$	2.79$^{(10)}$	0.47$^{(8)}$	0.47$^{(7)}$	0.45$^{(5)}$	0.47$^{(6)}$	0.45$^{(4)}$
S-MAR (S6)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.77$^{(9)}$	0.46$^{(3)}$	3.03$^{(10)}$	0.47$^{(6)}$	0.48$^{(7)}$	0.60$^{(8)}$	0.47$^{(5)}$	0.46$^{(4)}$
MNAR (S7)	**1.65**	1.70	2.88	2.01	4.01	1.99	1.99	2.07	1.98	1.97
S-MNAR (S8)	**1.58**	1.65	3.00	2.03	4.02	2.00	2.01	2.11	2.01	1.99
					40% Missingness					
MCAR (S2)	**0.05**$^{(1)}$	0.05$^{(2)}$	2.99$^{(9)}$	0.73$^{(3)}$	4.76$^{(10)}$	0.77$^{(8)}$	0.76$^{(6)}$	0.77$^{(7)}$	0.76$^{(5)}$	0.75$^{(4)}$
MAR (S3)	**0.04**$^{(1)}$	0.06$^{(2)}$	2.83$^{(9)}$	0.78$^{(4)}$	4.32$^{(10)}$	0.78$^{(7)}$	0.78$^{(6)}$	0.90$^{(8)}$	0.78$^{(5)}$	0.76$^{(3)}$
S-MAR (S4)	**0.06**$^{(1)}$	0.06$^{(2)}$	2.82$^{(9)}$	0.76$^{(3)}$	4.62$^{(10)}$	0.79$^{(8)}$	0.79$^{(7)}$	0.78$^{(5)}$	0.79$^{(6)}$	0.78$^{(4)}$
MAR (S5)	**0.02**$^{(1)}$	0.03$^{(2)}$	2.67$^{(9)}$	0.69$^{(3)}$	4.21$^{(10)}$	0.75$^{(7)}$	0.75$^{(6)}$	0.82$^{(8)}$	0.74$^{(5)}$	0.73$^{(4)}$
S-MAR (S6)	**0.04**$^{(1)}$	0.05$^{(2)}$	3.09$^{(9)}$	0.76$^{(3)}$	5.12$^{(10)}$	0.80$^{(6)}$	0.79$^{(5)}$	1.01$^{(8)}$	0.80$^{(7)}$	0.77$^{(4)}$
MNAR (S7)	8.58	**8.46**	10.91	9.14	12.07	9.28	9.17	10.55	9.14	9.14
S-MNAR (S8)	**8.34**	8.40	11.39	9.25	12.61	9.29	9.30	9.70	9.26	9.28
					60% Missingness					
MCAR (S2)	**0.08**$^{(1)}$	0.08$^{(2)}$	5.74$^{(9)}$	1.41$^{(3)}$	11.35$^{(10)}$	1.53$^{(7)}$	1.51$^{(6)}$	2.11$^{(8)}$	1.50$^{(5)}$	1.46$^{(4)}$
MAR (S3)	**0.08**$^{(1)}$	0.11$^{(2)}$	6.17$^{(9)}$	1.54$^{(6)}$	10.58$^{(10)}$	1.54$^{(7)}$	1.53$^{(5)}$	1.54$^{(8)}$	1.52$^{(4)}$	1.50$^{(3)}$
S-MAR (S4)	**0.07**$^{(1)}$	0.07$^{(2)}$	6.05$^{(9)}$	1.48$^{(3)}$	10.42$^{(10)}$	1.56$^{(8)}$	1.56$^{(7)}$	1.54$^{(6)}$	1.54$^{(5)}$	1.50$^{(4)}$
MAR (S5)	**0.04**$^{(1)}$	0.05$^{(2)}$	5.11$^{(9)}$	1.51$^{(3)}$	8.41$^{(10)}$	1.60$^{(7)}$	1.59$^{(6)}$	1.63$^{(8)}$	1.58$^{(5)}$	1.56$^{(4)}$
S-MAR (S6)	**0.05**$^{(1)}$	0.08$^{(2)}$	5.20$^{(9)}$	1.60$^{(3)}$	8.47$^{(10)}$	1.67$^{(7)}$	1.66$^{(6)}$	1.83$^{(8)}$	1.65$^{(5)}$	1.61$^{(4)}$
MNAR (S7)	28.40	**28.39**	37.63	31.77	38.07	31.97	32.02	32.37	31.91	31.94
S-MNAR (S8)	**27.86**	27.94	38.05	31.69	38.22	32.04	32.17	33.13	32.05	32.10

Table 5: *Summary statistics of squared bias using adjusted weigths, analysed under the M2 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(0.5,0.008)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	2.46	3.63	6.30	3.46	9.28	**2.38**	2.40	/	/	/
	20% Missingness									
MCAR (S2)	3.15$^{(6)}$	4.64$^{(8)}$	8.36$^{(9)}$	4.32$^{(7)}$	9.47$^{(10)}$	**3.03**$^{(1)}$	3.06$^{(4)}$	3.13$^{(5)}$	3.06$^{(3)}$	3.06$^{(2)}$
MAR (S3)	2.98$^{(6)}$	4.26$^{(8)}$	7.60$^{(9)}$	3.96$^{(7)}$	7.81$^{(10)}$	**2.87**$^{(1)}$	2.89$^{(4)}$	3.43$^{(5)}$	2.88$^{(3)}$	2.88$^{(2)}$
S-MAR (S4)	2.92$^{(6)}$	4.16$^{(8)}$	7.87$^{(9)}$	3.92$^{(7)}$	9.71$^{(10)}$	2.87$^{(2)}$	2.88$^{(4)}$	2.90$^{(5)}$	2.88$^{(3)}$	**2.87**$^{(1)}$
MAR (S5)	3.00$^{(6)}$	4.16$^{(8)}$	8.23$^{(9)}$	3.89$^{(7)}$	11.68$^{(10)}$	**2.91**$^{(1)}$	2.93$^{(3)}$	2.95$^{(5)}$	2.93$^{(4)}$	2.92$^{(2)}$
S-MAR (S6)	3.01$^{(5)}$	4.25$^{(8)}$	8.21$^{(9)}$	3.95$^{(7)}$	11.84$^{(10)}$	**2.90**$^{(1)}$	2.92$^{(3)}$	3.58$^{(6)}$	2.92$^{(4)}$	2.91$^{(2)}$
MNAR (S7)	4.70	6.11	8.93	5.71	12.77	**4.49**	4.51	5.01	4.50	4.50
S-MNAR (S8)	4.57	6.13	9.07	5.79	11.90	**4.46**	4.49	5.17	4.49	4.49
	40% Missingness									
MCAR (S2)	4.14$^{(6)}$	5.96$^{(8)}$	11.33$^{(9)}$	5.58$^{(7)}$	13.94$^{(10)}$	4.06$^{(4)}$	4.04$^{(2)}$	4.05$^{(3)}$	4.06$^{(5)}$	**4.03**$^{(1)}$
MAR (S3)	3.94$^{(2)}$	5.38$^{(8)}$	11.10$^{(9)}$	5.18$^{(7)}$	12.78$^{(10)}$	3.96$^{(4)}$	3.97$^{(5)}$	4.48$^{(6)}$	3.96$^{(3)}$	**3.90**$^{(1)}$
S-MAR (S4)	3.87$^{(3)}$	5.24$^{(8)}$	11.36$^{(9)}$	5.04$^{(7)}$	14.67$^{(10)}$	3.91$^{(4)}$	3.92$^{(6)}$	3.87$^{(2)}$	3.92$^{(5)}$	**3.86**$^{(1)}$
MAR (S5)	4.05$^{(5)}$	5.33$^{(8)}$	11.08$^{(9)}$	5.10$^{(7)}$	13.37$^{(10)}$	4.03$^{(2)}$	4.04$^{(4)}$	4.19$^{(6)}$	4.03$^{(3)}$	**4.00**$^{(1)}$
S-MAR (S6)	4.01$^{(3)}$	5.43$^{(8)}$	11.08$^{(9)}$	5.17$^{(7)}$	15.39$^{(10)}$	4.01$^{(2)}$	4.04$^{(5)}$	4.71$^{(6)}$	4.04$^{(4)}$	**3.99**$^{(1)}$
MNAR (S7)	13.00	14.61	18.52	14.14	20.76	12.78	12.64	16.97	12.61	**12.59**
S-MNAR (S8)	**12.58**	14.55	19.17	14.32	20.74	12.69	12.70	14.37	12.68	12.65
	60% Missingness									
MCAR (S2)	6.40$^{(5)}$	8.64$^{(8)}$	17.17$^{(9)}$	7.79$^{(6)}$	27.29$^{(10)}$	6.08$^{(2)}$	6.10$^{(3)}$	8.04$^{(7)}$	6.11$^{(4)}$	**6.06**$^{(1)}$
MAR (S3)	6.25$^{(6)}$	8.03$^{(8)}$	17.50$^{(9)}$	7.44$^{(7)}$	25.10$^{(10)}$	6.03$^{(3)}$	6.02$^{(4)}$	6.10$^{(5)}$	6.02$^{(2)}$	**6.02**$^{(1)}$
S-MAR (S4)	6.26$^{(6)}$	7.8$^{(8)}$3	17.51$^{(9)}$	7.35$^{(7)}$	25.19$^{(10)}$	6.11$^{(4)}$	6.10$^{(3)}$	6.15$^{(5)}$	6.10$^{(2)}$	**6.02**$^{(1)}$
MAR (S5)	5.95$^{(4)}$	7.52$^{(8)}$	15.39$^{(9)}$	7.14$^{(7)}$	20.43$^{(10)}$	5.95$^{(5)}$	5.92$^{(2)}$	6.01$^{(6)}$	5.93$^{(3)}$	**5.87**$^{(1)}$
S-MAR (S6)	6.14$^{(5)}$	7.97$^{(8)}$	15.76$^{(9)}$	7.48$^{(7)}$	20.52$^{(10)}$	6.09$^{(2)}$	6.11$^{(4)}$	6.72$^{(6)}$	6.10$^{(3)}$	**6.07**$^{(1)}$
MNAR (S7)	**36.43**	38.84	48.25	39.63	50.18	37.77	37.71	38.60	37.66	37.73
S-MNAR (S8)	**36.09**	38.77	48.00	39.13	50.29	37.38	37.52	39.84	37.42	37.48

Table 6: *Summary statistics of MSE using adjusted weigths, analysed under the M2 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(0.5,0.008)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	0.92	0.89	0.88	0.90	0.85	0.95	0.95	/	/	/
	20% Missingness									
MCAR (S2)	0.90	0.87	0.86	0.89	0.82	0.94	0.94	0.94	0.94	0.94
MAR (S3)	0.91	0.87	0.88	0.90	0.83	0.95	0.95	0.94	0.95	0.95
S-MAR (S4)	0.91	0.87	0.88	0.91	0.84	0.95	0.95	0.95	0.95	0.95
MAR (S5)	0.91	0.87	0.87	0.90	0.82	0.95	0.94	0.95	0.94	0.95
S-MAR (S6)	0.91	0.88	0.87	0.91	0.80	0.95	0.95	0.93	0.95	0.95
MNAR (S7)	0.81	0.81	0.78	0.80	0.73	0.84	0.84	0.83	0.84	0.84
S-MNAR (S8)	0.82	0.81	0.78	0.80	0.74	0.84	0.84	0.83	0.84	0.85
	40% Missingness									
MCAR (S2)	0.89	0.84	0.85	0.90	0.80	0.95	0.94	0.95	0.94	0.95
MAR(S3)	0.89	0.86	0.87	0.91	0.81	0.95	0.95	0.94	0.95	0.95
S-MAR (S4)	0.89	0.86	0.86	0.91	0.80	0.95	0.95	0.95	0.95	0.95
MAR (S5)	0.89	0.86	0.88	0.91	0.82	0.94	0.94	0.94	0.94	0.95
S-MAR (S6)	0.89	0.86	0.85	0.91	0.76	0.95	0.95	0.93	0.95	0.95
MNAR (S7)	0.56	0.62	0.54	0.58	0.50	0.59	0.59	0.56	0.59	0.60
S-MNAR (S8)	0.60	0.64	0.57	0.59	0.53	0.60	0.60	0.59	0.61	0.61
	60% Missingness									
MCAR (S2)	0.86	0.81	0.81	0.91	0.69	0.95	0.95	0.92	0.95	0.96
MAR (S3)	0.86	0.82	0.83	0.91	0.70	0.95	0.95	0.95	0.95	0.95
S-MAR (S4)	0.86	0.82	0.81	0.91	0.70	0.94	0.94	0.94	0.94	0.94
MAR (S5)	0.87	0.83	0.83	0.92	0.71	0.95	0.95	0.95	0.95	0.95
S-MAR (S6)	0.86	0.83	0.83	0.91	0.73	0.95	0.95	0.94	0.95	0.95
MNAR (S7)	0.38	0.43	0.33	0.39	0.32	0.38	0.39	0.38	0.39	0.39
S-MNAR (S8)	0.41	0.45	0.35	0.41	0.34	0.40	0.40	0.39	0.40	0.40

Table 7: *Nominal coverage probabilities using adjusted weigths, analysed under the M2 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(0.5,0.008)-distribution.*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	16.89	0.65	0.63	0.42	0.48	**0.18**	0.22	/	/	/
					20% Missingness					
MCAR (S2)	16.90$^{(10)}$	0.88$^{(9)}$	0.85$^{(8)}$	0.56$^{(6)}$	0.64$^{(7)}$	**0.34**$^{(1)}$	0.38$^{(4)}$	0.37$^{(2)}$	0.38$^{(3)}$	0.51$^{(5)}$
MAR (S5)	10.96$^{(10)}$	0.49$^{(9)}$	0.47$^{(8)}$	0.31$^{(6)}$	0.36$^{(7)}$	**0.19**$^{(1)}$	0.24$^{(2)}$	0.26$^{(4)}$	0.25$^{(3)}$	0.31$^{(5)}$
S-MAR (S6)	11.25$^{(10)}$	0.54$^{(9)}$	0.52$^{(8)}$	0.35$^{(6)}$	0.40$^{(7)}$	**0.21**$^{(1)}$	0.26$^{(2)}$	0.32$^{(4)}$	0.27$^{(3)}$	0.34$^{(5)}$
MNAR (S7)	7.80	0.60	0.58	0.54	**0.53**	0.70	0.82	0.55	0.63	0.56
S-MNAR (S8)	8.01	0.59	0.58	0.53	**0.52**	0.67	0.79	0.55	0.61	0.53
					40% Missingness					
MCAR (S2)	16.70$^{(10)}$	1.16$^{(9)}$	1.10$^{(8)}$	0.73$^{(5)}$	0.82$^{(7)}$	**0.58**$^{(1)}$	0.64$^{(2)}$	0.68$^{(4)}$	0.67$^{(3)}$	0.79$^{(6)}$
MAR (S5)	4.76$^{(10)}$	0.23$^{(9)}$	0.22$^{(8)}$	**0.14**$^{(1)}$	0.18$^{(5)}$	0.18$^{(2)}$	0.22$^{(7)}$	0.22$^{(6)}$	0.18$^{(3)}$	0.18$^{(4)}$
S-MAR (S6)	5.12$^{(10)}$	0.25$^{(8)}$	0.25$^{(7)}$	**0.18**$^{(1)}$	0.20$^{(2)}$	0.22$^{(4)}$	0.27$^{(9)}$	0.23$^{(6)}$	0.22$^{(3)}$	0.22$^{(5)}$
MNAR (S7)	**0.88**	4.27	4.24	4.42	4.21	4.81	4.96	4.05	4.46	4.27
S-MNAR (S8)	**1.45**	4.16	4.07	4.13	3.92	4.51	4.66	3.84	4.14	3.97
					60% Missingness					
MCAR (S2)	16.36$^{(10)}$	1.72$^{(9)}$	1.52$^{(8)}$	**1.11**$^{(1)}$	1.19$^{(2)}$	1.20$^{(3)}$	1.28$^{(4)}$	1.45$^{(7)}$	1.30$^{(5)}$	1.41$^{(6)}$
MAR (S5)	0.38$^{(10)}$	0.18$^{(9)}$	0.16$^{(7)}$	0.10$^{(2)}$	0.12$^{(4)}$	0.15$^{(6)}$	0.17$^{(8)}$	**0.08**$^{(1)}$	0.12$^{(3)}$	0.12$^{(5)}$
S-MAR (S6)	0.60$^{(10)}$	0.15$^{(7)}$	0.13$^{(4)}$	**0.09**$^{(1)}$	0.10$^{(2)}$	0.22$^{(8)}$	0.25$^{(9)}$	0.14$^{(6)}$	0.14$^{(5)}$	0.12$^{(3)}$
MNAR (S7)	**5.60**	19.68	19.47	19.42	18.54	19.20	19.21	17.99	19.33	19.01
S-MNAR (S8)	**6.23**	19.00	18.60	17.97	17.16	17.83	17.96	16.85	17.92	17.63

Table 8: *Summary statistics of squared bias using adjusted weigths, analysed under the M1 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(2,1)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	**0.04**	0.04	0.86	0.36	1.74	0.34	0.34	/	/	/
					20% Missingness					
MCAR (S2)	**0.03**$^{(1)}$	0.04$^{(2)}$	1.78$^{(9)}$	0.54$^{(7)}$	2.54$^{(10)}$	0.52$^{(4)}$	0.52$^{(5)}$	0.71$^{(8)}$	0.53$^{(6)}$	0.46$^{(3)}$
MAR (S5)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.46$^{(9)}$	0.50$^{(7)}$	2.05$^{(10)}$	0.50$^{(4)}$	0.50$^{(6)}$	0.50$^{(8)}$	0.50$^{(5)}$	0.43$^{(3)}$
S-MAR (S6)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.51$^{(9)}$	0.54$^{(7)}$	2.44$^{(10)}$	0.53$^{(4)}$	0.54$^{(6)}$	0.55$^{(8)}$	0.53$^{(5)}$	0.46$^{(3)}$
MNAR (S7)	**1.65**	1.69	2.90	2.02	4.02	2.01	2.00	2.08	2.00	1.94
S-MNAR (S8)	**1.59**	1.66	3.08	2.06	4.06	2.02	2.03	2.08	2.03	1.95
					40% Missingness					
MCAR (S2)	**0.05**$^{(1)}$	0.05$^{(2)}$	3.27$^{(9)}$	0.79$^{(4)}$	4.54$^{(10)}$	0.82$^{(6)}$	0.83$^{(7)}$	0.82$^{(5)}$	0.84$^{(8)}$	0.73$^{(3)}$
MAR (S5)	**0.04**$^{(1)}$	0.06$^{(2)}$	3.03$^{(9)}$	0.84$^{(4)}$	4.10$^{(10)}$	0.84$^{(5)}$	0.85$^{(7)}$	0.96$^{(8)}$	0.85$^{(6)}$	0.74$^{(3)}$
S-MAR (S6)	**0.06**$^{(1)}$	0.06$^{(2)}$	2.93$^{(9)}$	0.82$^{(4)}$	4.49$^{(10)}$	0.87$^{(5)}$	0.88$^{(7)}$	0.88$^{(8)}$	0.88$^{(6)}$	0.77$^{(3)}$
MNAR (S7)	8.55	**8.42**	10.81	9.17	11.94	9.30	9.22	9.77	9.18	9.09
S-MNAR (S8)	**8.36**	8.42	11.39	9.36	12.53	9.38	9.38	9.70	9.33	9.21
					60% Missingness					
MCAR (S2)	**0.08**$^{(1)}$	0.08$^{(2)}$	5.55$^{(9)}$	1.54$^{(4)}$	10.64$^{(10)}$	1.67$^{(5)}$	1.67$^{(8)}$	1.77$^{(8)}$	1.67$^{(6)}$	1.46$^{(3)}$
MAR (S5)	**0.08**$^{(1)}$	0.11$^{(2)}$	6.16$^{(9)}$	1.69$^{(7)}$	10.03$^{(10)}$	1.68$^{(7)}$	1.69$^{(8)}$	1.67$^{(4)}$	1.68$^{(6)}$	1.46$^{(3)}$
S-MAR (S6)	**0.07**$^{(1)}$	0.07$^{(2)}$	6.00$^{(9)}$	1.62$^{(4)}$	9.96$^{(10)}$	1.69$^{(5)}$	1.70$^{(7)}$	1.73$^{(8)}$	1.70$^{(6)}$	1.47$^{(3)}$
MNAR (S7)	28.40	**28.39**	35.42	32.04	37.73	32.25	32.24	32.85	32.13	31.96
S-MNAR (S8)	**27.86**	27.94	35.89	31.94	37.94	32.20	32.37	32.50	32.27	32.10

Table 9: *Summary statistics of squared bias using adjusted weigths, analysed under the M2 simulation mechanism. The prior distribution of the precision parameters* σ_u^{-2} *and* σ_v^{-2} *are assumed to follow a Gamma(2,1)-distribution.* ($\times 10^3$)

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	20.61	6.22	5.53	2.74	2.76	2.01	**1.99**	/	/	/
	20% Missingness									
MCAR (S2)	21.71$^{(10)}$	8.01$^{(9)}$	6.93$^{(8)}$	3.29$^{(7)}$	3.29$^{(6)}$	2.46$^{(4)}$	2.43$^{(3)}$	**2.40**$^{(1)}$	2.41$^{(2)}$	2.67$^{(5)}$
MAR (S5)	15.80$^{(10)}$	7.04$^{(9)}$	5.98$^{(8)}$	2.74$^{(7)}$	2.67$^{(6)}$	2.34$^{(4)}$	**2.26**$^{(1)}$	2.27$^{(3)}$	2.27$^{(2)}$	2.40$^{(5)}$
S-MAR (S6)	16.07$^{(10)}$	6.84$^{(9)}$	5.84$^{(8)}$	2.70$^{(7)}$	2.68$^{(6)}$	2.26$^{(2)}$	**2.23**$^{(1)}$	2.40$^{(4)}$	2.26$^{(3)}$	2.38$^{(5)}$
MNAR (S7)	12.76	7.08	6.10	2.98	2.87	2.77	2.81	**2.56**	2.63	2.64
S-MNAR (S8)	12.96	7.25	6.27	3.05	2.97	2.79	2.83	**2.61**	2.62	2.64
	40% Missingness									
MCAR (S2)	23.05$^{(10)}$	10.49$^{(9)}$	8.67$^{(8)}$	3.87$^{(7)}$	3.85$^{(6)}$	3.01$^{(2)}$	3.03$^{(3)}$	3.08$^{(4)}$	**2.99**$^{(1)}$	3.30$^{(5)}$
MAR (S5)	11.74$^{(10)}$	8.98$^{(9)}$	7.16$^{(8)}$	2.88$^{(7)}$	2.70$^{(5)}$	2.77$^{(6)}$	2.58$^{(3)}$	**2.48**$^{(1)}$	2.50$^{(2)}$	2.61$^{(4)}$
S-MAR (S6)	11.64$^{(10)}$	8.24$^{(9)}$	6.64$^{(8)}$	2.69$^{(7)}$	2.63$^{(5)}$	2.65$^{(6)}$	2.51$^{(4)}$	2.42$^{(2)}$	**2.40**$^{(1)}$	2.49$^{(3)}$
MNAR (S7)	7.91	12.62	10.95	7.09	6.68	7.13	7.19	**6.35**	6.72	6.66
S-MNAR (S8)	8.53	12.69	10.94	6.88	6.51	6.97	7.03	**6.26**	6.43	6.40
	60% Missingness									
MCAR (S2)	26.09$^{(10)}$	15.47$^{(9)}$	11.08$^{(8)}$	4.84$^{(7)}$	4.51$^{(6)}$	**4.05**$^{(1)}$	4.09$^{(3)}$	4.40$^{(4)}$	4.08$^{(2)}$	4.46$^{(5)}$
MAR (S5)	11.44$^{(10)}$	13.01$^{(9)}$	9.01$^{(8)}$	3.25$^{(7)}$	2.93$^{(3)}$	3.19$^{(6)}$	3.00$^{(5)}$	**2.92**$^{(1)}$	2.80$^{(2)}$	2.99$^{(4)}$
S-MAR (S6)	11.92$^{(10)}$	12.91$^{(9)}$	8.57$^{(8)}$	3.14$^{(7)}$	**2.79**$^{(1)}$	3.20$^{(6)}$	3.08$^{(5)}$	3.03$^{(3)}$	2.87$^{(2)}$	3.03$^{(4)}$
MNAR (S7)	**16.72**	31.75	27.97	22.38	21.08	21.91	21.83	20.83	21.85	21.75
S-MNAR (S8)	**18.11**	31.63	26.82	21.05	19.74	20.74	20.78	19.51	20.62	20.55

Table 10: *Summary statistics of MSE using adjusted weigths, analysed under the M1 simulation mechanism. The prior distribution of the precision parameters* σ_u^{-2} *and* σ_v^{-2} *are assumed to follow a Gamma(2,1)-distribution. (*$\times 10^3$*)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	2.46	3.63	6.65	3.45	9.11	**2.40**	2.42	/	/	/
	20% Missingness									
MCAR (S2)	$3.16^{(5)}$	$4.65^{(8)}$	$9.12^{(9)}$	$4.39^{(7)}$	$9.38^{(10)}$	$3.08^{(2)}$	$3.11^{(4)}$	$3.94^{(6)}$	$3.11^{(3)}$	$\mathbf{3.08}^{(1)}$
MAR (S5)	$2.98^{(6)}$	$4.27^{(8)}$	$8.42^{(10)}$	$4.02^{(7)}$	$7.84^{(9)}$	$2.92^{(2)}$	$2.94^{(5)}$	$2.93^{(4)}$	$2.93^{(3)}$	$\mathbf{2.90}^{(1)}$
S-MAR (S6)	$2.92^{(4)}$	$4.15^{(8)}$	$8.38^{(9)}$	$3.96^{(7)}$	$9.65^{(10)}$	$2.90^{(2)}$	$2.92^{(3)}$	$2.95^{(6)}$	$2.93^{(5)}$	$\mathbf{2.88}^{(1)}$
MNAR (S7)	4.70	6.08	9.50	5.72	12.67	4.52	4.54	5.00	4.54	**4.51**
S-MNAR (S8)	4.59	6.13	9.81	5.82	11.88	4.49	4.52	4.84	4.52	**4.48**
	40% Missingness									
MCAR (S2)	$4.15^{(6)}$	$5.96^{(8)}$	$12.64^{(9)}$	$5.48^{(7)}$	$13.64^{(10)}$	$3.99^{(2)}$	$4.01^{(3)}$	$4.02^{(5)}$	$4.01^{(4)}$	$\mathbf{3.97}^{(1)}$
MAR (S5)	$3.95^{(5)}$	$5.38^{(8)}$	$12.27^{(9)}$	$5.07^{(7)}$	$12.35^{(10)}$	$3.89^{(2)}$	$3.92^{(4)}$	$4.43^{(6)}$	$3.91^{(3)}$	$\mathbf{3.85}^{(1)}$
S-MAR (S6)	$3.86^{(5)}$	$5.20^{(8)}$	$12.19^{(9)}$	$4.90^{(7)}$	$14.22^{(10)}$	$3.82^{(2)}$	$3.85^{(3)}$	$3.87^{(6)}$	$3.85^{(4)}$	$\mathbf{3.80}^{(1)}$
MNAR (S7)	12.98	14.55	19.20	14.05	20.38	12.73	12.68	14.41	12.64	**12.61**
S-MNAR (S8)	**12.60**	14.55	20.10	14.23	20.33	12.68	12.72	13.75	12.68	12.63
	60% Missingness									
MCAR (S2)	$6.40^{(6)}$	$8.64^{(8)}$	$18.08^{(9)}$	$7.73^{(7)}$	$26.19^{(10)}$	$6.07^{(2)}$	$6.09^{(3)}$	$6.34^{(5)}$	$6.10^{(4)}$	$\mathbf{6.01}^{(1)}$
MAR (S5)	$6.25^{(6)}$	$8.03^{(8)}$	$18.54^{(9)}$	$7.39^{(7)}$	$24.13^{(10)}$	$6.03^{(2)}$	$6.04^{(3)}$	$6.05^{(5)}$	$6.05^{(4)}$	$\mathbf{5.97}^{(1)}$
S-MAR (S6)	$6.26^{(6)}$	$7.83^{(8)}$	$18.55^{(9)}$	$7.32^{(7)}$	$24.36^{(10)}$	$6.10^{(2)}$	$6.11^{(3)}$	$6.25^{(5)}$	$6.11^{(4)}$	$\mathbf{6.03}^{(1)}$
MNAR (S7)	**36.43**	38.84	47.57	39.46	49.21	37.78	37.80	38.98	37.72	37.75
S-MNAR (S8)	**36.09**	38.77	47.16	39.14	49.58	37.47	37.67	38.30	37.59	37.64

Table 11: *Summary statistics of MSE using adjusted weigths, analysed under the M2 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(2,1)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	16.86	0.65	0.62	0.39	0.46	**0.18**	0.21	/	/	/
					20% Missingness					
MCAR (S2)	$16.90^{(10)}$	$0.88^{(9)}$	$0.84^{(8)}$	$0.53^{(6)}$	$0.61^{(7)}$	$0.48^{(4)}$	$0.38^{(2)}$	$\mathbf{0.37}^{(1)}$	$0.38^{(3)}$	$0.50^{(5)}$
MAR (S5)	$10.96^{(10)}$	$0.49^{(9)}$	$0.47^{(8)}$	$0.29^{(5)}$	$0.35^{(7)}$	$\mathbf{0.13}^{(1)}$	$0.24^{(3)}$	$0.23^{(2)}$	$0.24^{(4)}$	$0.31^{(6)}$
S-MAR (S6)	$11.25^{(10)}$	$0.54^{(9)}$	$0.51^{(8)}$	$0.33^{(5)}$	$0.39^{(7)}$	$\mathbf{0.15}^{(1)}$	$0.26^{(2)}$	$0.29^{(4)}$	$0.27^{(3)}$	$0.34^{(6)}$
MNAR (S7)	7.80	0.60	0.57	0.54	0.52	0.66	0.80	**0.50**	0.62	0.56
S-MNAR (S8)	8.01	0.59	0.57	0.53	**0.51**	0.61	0.76	0.61	0.59	0.53
					40% Missingness					
MCAR (S2)	$16.70^{(10)}$	$1.16^{(9)}$	$1.09^{(8)}$	$0.69^{(4)}$	$0.79^{(6)}$	$1.01^{(7)}$	$\mathbf{0.64}^{(1)}$	$0.66^{(2)}$	$0.66^{(3)}$	$0.77^{(5)}$
MAR (S5)	$4.76^{(10)}$	$0.23^{(9)}$	$0.22^{(8)}$	$\mathbf{0.14}^{(1)}$	$0.18^{(5)}$	$0.16^{(2)}$	$0.22^{(7)}$	$0.21^{(6)}$	$0.18^{(3)}$	$0.18^{(4)}$
S-MAR (S6)	$5.12^{(10)}$	$0.25^{(8)}$	$0.25^{(7)}$	$\mathbf{0.18}^{(1)}$	$0.20^{(3)}$	$0.18^{(2)}$	$0.27^{(9)}$	$0.22^{(4)}$	$0.22^{(5)}$	$0.22^{(6)}$
MNAR (S7)	**0.88**	4.27	4.22	4.46	4.23	4.57	4.92	4.05	4.45	4.27
S-MNAR (S8)	**1.45**	4.16	4.04	4.15	3.91	4.16	4.58	3.69	4.12	3.96
					60% Missingness					
MCAR (S2)	$16.36^{(10)}$	$1.72^{(8)}$	$1.53^{(7)}$	$\mathbf{1.07}^{(1)}$	$1.17^{(2)}$	$2.11^{(9)}$	$1.27^{(3)}$	$1.40^{(6)}$	$1.28^{(4)}$	$1.38^{(5)}$
MAR (S5)	$0.38^{(10)}$	$0.18^{(8)}$	$0.14^{(6)}$	$0.10^{(2)}$	$0.12^{(5)}$	$0.26^{(9)}$	$0.17^{(7)}$	$\mathbf{0.07}^{(1)}$	$0.11^{(3)}$	$0.11^{(4)}$
S-MAR (S6)	$0.60^{(10)}$	$0.15^{(7)}$	$0.12^{(4)}$	$\mathbf{0.09}^{(1)}$	$0.09^{(2)}$	$0.27^{(9)}$	$0.24^{(8)}$	$0.13^{(5)}$	$0.13^{(6)}$	$0.11^{(3)}$
MNAR (S7)	**5.60**	19.68	19.38	19.51	18.57	18.50	19.22	18.20	19.34	19.03
S-MNAR (S8)	**6.23**	19.00	18.43	17.96	17.09	16.68	17.81	16.23	17.84	17.59

Table 12: *Summary statistics of squared bias using adjusted weigths, analysed under the M1 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(1,0.5)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	**0.04**	0.04	0.86	0.33	1.68	0.32	0.32	/	/	/
	20% Missingness									
MCAR (S2)	**0.03**$^{(1)}$	0.04$^{(2)}$	1.74$^{(9)}$	0.51$^{(7)}$	2.46$^{(10)}$	0.49$^{(5)}$	0.49$^{(6)}$	0.66$^{(8)}$	0.49$^{(4)}$	0.44$^{(3)}$
MAR (S5)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.43$^{(9)}$	0.46$^{(5)}$	1.97$^{(10)}$	0.46$^{(4)}$	0.47$^{(7)}$	0.53$^{(8)}$	0.47$^{(6)}$	0.41$^{(3)}$
S-MAR (S6)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.49$^{(9)}$	0.51$^{(7)}$	2.36$^{(10)}$	0.50$^{(5)}$	0.50$^{(6)}$	0.56$^{(8)}$	0.50$^{(4)}$	0.45$^{(3)}$
MNAR (S7)	**1.65**	1.69	2.88	2.00	3.93	1.98	1.97	2.16	1.98	1.92
S-MNAR (S8)	**1.59**	1.66	3.05	2.03	3.97	1.99	2.00	1.99	2.00	1.94
	40% Missingness									
MCAR (S2)	**0.05**$^{(1)}$	0.05$^{(2)}$	3.19$^{(9)}$	0.74$^{(4)}$	4.39$^{(10)}$	0.77$^{(5)}$	0.78$^{(6)}$	0.90$^{(8)}$	0.78$^{(7)}$	0.70$^{(3)}$
MAR (S5)	**0.04**$^{(1)}$	0.06$^{(2)}$	2.96$^{(9)}$	0.79$^{(4)}$	3.95$^{(10)}$	0.79$^{(5)}$	0.79$^{(6)}$	1.08$^{(8)}$	0.80$^{(7)}$	0.71$^{(3)}$
S-MAR (S6)	**0.06**$^{(1)}$	0.06$^{(2)}$	2.89$^{(9)}$	0.77$^{(4)}$	4.35$^{(10)}$	0.81$^{(5)}$	0.82$^{(6)}$	0.84$^{(8)}$	0.83$^{(7)}$	0.73$^{(3)}$
MNAR (S7)	8.55	**8.42**	10.81	9.11	11.76	9.24	9.15	9.63	9.14	9.05
S-MNAR (S8)	**8.36**	8.42	11.40	9.30	12.37	9.31	9.32	9.56	9.27	9.17
	60% Missingness									
MCAR (S2)	**0.08**$^{(1)}$	0.08$^{(2)}$	5.51$^{(9)}$	1.45$^{(4)}$	10.34$^{(10)}$	1.56$^{(6)}$	1.57$^{(7)}$	1.60$^{(8)}$	1.56$^{(5)}$	1.39$^{(3)}$
MAR (S5)	**0.08**$^{(1)}$	0.11$^{(2)}$	6.09$^{(9)}$	1.58$^{(7)}$	9.73$^{(10)}$	1.57$^{(6)}$	1.58$^{(8)}$	1.56$^{(5)}$	1.56$^{(4)}$	1.40$^{(3)}$
S-MAR (S6)	**0.07**$^{(1)}$	0.07$^{(2)}$	5.92$^{(9)}$	1.52$^{(4)}$	9.69$^{(10)}$	1.58$^{(5)}$	1.59$^{(8)}$	1.59$^{(7)}$	1.59$^{(6)}$	1.41$^{(3)}$
MNAR (S7)	28.40	**28.39**	35.98	31.76	37.18	31.96	31.95	32.35	31.86	31.75
S-MNAR (S8)	**27.86**	27.94	36.42	31.68	37.44	31.91	32.09	33.18	32.02	31.85

Table 13: *Summary statistics of squared bias using adjusted weigths, analysed under the M2 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(1,0.5)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	20.58	6.22	5.15	2.25	2.28	1.70	**1.66**	/	/	/
					20% Missingness					
MCAR (S2)	21.71$^{(10)}$	8.01$^{(9)}$	6.39$^{(8)}$	2.73$^{(6)}$	2.75$^{(7)}$	2.25$^{(4)}$	2.04$^{(2)}$	2.08$^{(3)}$	**2.03**$^{(1)}$	2.37$^{(5)}$
MAR (S5)	15.80$^{(10)}$	7.04$^{(9)}$	5.48$^{(8)}$	2.22$^{(7)}$	2.18$^{(6)}$	1.98$^{(4)}$	**1.86**$^{(1)}$	1.89$^{(2)}$	1.93$^{(3)}$	2.13$^{(5)}$
S-MAR (S6)	16.07$^{(10)}$	6.84$^{(9)}$	5.36$^{(8)}$	2.20$^{(7)}$	2.20$^{(6)}$	1.88$^{(2)}$	**1.84**$^{(1)}$	1.94$^{(4)}$	1.92$^{(3)}$	2.11$^{(5)}$
MNAR (S7)	12.76	7.08	5.59	2.48	2.38	2.43	2.40	**2.19**	2.26	2.36
S-MNAR (S8)	12.96	7.25	5.74	2.53	2.48	2.38	2.41	2.31	**2.24**	2.36
					40% Missingness					
MCAR (S2)	23.05$^{(10)}$	10.49$^{(9)}$	7.89$^{(8)}$	3.23$^{(6)}$	3.25$^{(7)}$	3.06$^{(5)}$	2.56$^{(3)}$	2.56$^{(2)}$	**2.55**$^{(1)}$	2.93$^{(4)}$
MAR (S5)	11.74$^{(10)}$	8.98$^{(9)}$	6.37$^{(8)}$	2.29$^{(6)}$	2.15$^{(4)}$	2.44$^{(7)}$	**2.09**$^{(1)}$	2.15$^{(3)}$	2.12$^{(2)}$	2.29$^{(5)}$
S-MAR (S6)	11.64$^{(10)}$	8.24$^{(9)}$	5.93$^{(8)}$	2.15$^{(4)}$	2.12$^{(3)}$	2.32$^{(7)}$	2.06$^{(7)}$	2.17$^{(5)}$	**2.06**$^{(1)}$	2.19$^{(6)}$
MNAR (S7)	7.91	12.62	10.17	6.57	6.16	6.56	6.70	**6.07**	6.30	6.33
S-MNAR (S8)	8.53	12.69	10.14	6.32	5.96	6.20	6.48	**5.56**	6.00	6.06
					60% Missingness					
MCAR (S2)	26.09$^{(10)}$	15.47$^{(9)}$	9.88$^{(8)}$	4.10$^{(6)}$	3.85$^{(4)}$	4.48$^{(7)}$	**3.53**$^{(1)}$	3.73$^{(3)}$	3.55$^{(2)}$	4.00$^{(5)}$
MAR (S5)	11.44$^{(10)}$	13.01$^{(9)}$	7.78$^{(8)}$	2.57$^{(5)}$	2.34$^{(2)}$	2.89$^{(7)}$	**2.42**$^{(1)}$	2.43$^{(4)}$	2.35$^{(3)}$	2.58$^{(6)}$
S-MAR (S6)	11.92$^{(10)}$	12.91$^{(9)}$	7.40$^{(8)}$	2.48$^{(3)}$	**2.19**$^{(1)}$	2.81$^{(7)}$	2.49$^{(4)}$	2.56$^{(5)}$	2.40$^{(2)}$	2.62$^{(6)}$
MNAR (S7)	**16.72**	31.75	26.65	21.82	20.54	20.73	21.31	20.46	21.41	21.37
S-MNAR (S8)	**18.11**	31.63	25.49	20.40	19.12	19.12	20.10	18.45	20.10	20.11

Table 14: *Summary statistics of MSE using adjusted weigths, analysed under the M1 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(1,0.5)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	2.46	3.63	6.51	3.42	9.04	**2.38**	2.40	/	/	/
					20% Missingness					
MCAR (S2)	3.16$^{(5)}$	4.65$^{(8)}$	8.87$^{(9)}$	4.35$^{(7)}$	9.31$^{(10)}$	3.05$^{(2)}$	3.08$^{(4)}$	3.93$^{(6)}$	3.08$^{(4)}$	**3.05**$^{(1)}$
MAR (S5)	2.98$^{(5)}$	4.27$^{(8)}$	8.18$^{(10)}$	3.98$^{(7)}$	7.74$^{(9)}$	2.88$^{(2)}$	2.91$^{(4)}$	3.25$^{(6)}$	2.91$^{(4)}$	**2.87**$^{(1)}$
S-MAR (S6)	2.92$^{(5)}$	4.15$^{(8)}$	8.17$^{(9)}$	3.94$^{(7)}$	9.57$^{(10)}$	2.88$^{(2)}$	2.90$^{(3)}$	3.16$^{(6)}$	2.90$^{(4)}$	**2.86**$^{(1)}$
MNAR (S7)	4.70	6.08	9.28	5.68	12.58	4.50	4.52	5.01	4.51	**4.48**
S-MNAR (S8)	4.59	6.13	9.56	5.79	11.78	4.47	4.50	4.47	4.50	**4.46**
					40% Missingness					
MCAR (S2)	4.15$^{(5)}$	5.96$^{(8)}$	12.25$^{(9)}$	5.42$^{(7)}$	13.50$^{(10)}$	3.94$^{(2)}$	3.97$^{(4)}$	4.40$^{(6)}$	3.96$^{(3)}$	**3.93**$^{(1)}$
MAR (S5)	3.95$^{(5)}$	5.38$^{(8)}$	11.89$^{(9)}$	5.01$^{(6)}$	12.22$^{(10)}$	3.85$^{(2)}$	3.87$^{(4)}$	5.18$^{(7)}$	3.86$^{(3)}$	**3.80**$^{(1)}$
S-MAR (S6)	3.86$^{(6)}$	5.20$^{(8)}$	11.85$^{(9)}$	4.85$^{(7)}$	14.11$^{(10)}$	3.77$^{(2)}$	3.80$^{(3)}$	3.82$^{(5)}$	3.81$^{(4)}$	**3.75**$^{(1)}$
MNAR (S7)	12.98	14.55	18.93	13.97	20.17	12.68	12.62	14.14	12.60	**12.54**
S-MNAR (S8)	12.60	14.55	19.79	14.17	20.12	12.63	12.66	13.53	12.63	**12.58**
					60% Missingness					
MCAR (S2)	6.40$^{(6)}$	8.64$^{(8)}$	17.55$^{(9)}$	7.63$^{(7)}$	26.00$^{(10)}$	5.96$^{(2)}$	6.00$^{(4)}$	6.05$^{(5)}$	6.00$^{(3)}$	**5.91**$^{(1)}$
MAR (S5)	6.25$^{(6)}$	8.03$^{(8)}$	18.08$^{(9)}$	7.27$^{(7)}$	23.89$^{(10)}$	5.92$^{(2)}$	5.95$^{(4)}$	5.97$^{(5)}$	5.95$^{(3)}$	**5.85**$^{(1)}$
S-MAR (S6)	6.26$^{(6)}$	7.83$^{(8)}$	18.07$^{(9)}$	7.20$^{(7)}$	24.16$^{(10)}$	6.00$^{(2)}$	6.02$^{(5)}$	6.01$^{(3)}$	6.02$^{(4)}$	**5.91**$^{(1)}$
MNAR (S7)	**36.43**	38.84	47.63	39.19	48.62	37.51	37.51	38.33	37.46	37.47
S-MNAR (S8)	**36.09**	38.77	47.33	38.91	49.08	37.21	37.44	39.80	37.39	37.34

Table 15: *Summary statistics of MSE using adjusted weigths, analysed under the M2 simulation mechanism. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(1,0.5)-distribution.*($\times 10^3$)

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	20.61	6.22	1.34	**0.55**	0.61	0.80	0.74	/	/	/
					20% Missingness					
MCAR (S2)	21.71$^{(10)}$	8.01$^{(9)}$	1.71$^{(8)}$	**0.74**$^{(1)}$	0.82$^{(2)}$	1.01$^{(6)}$	0.93$^{(4)}$	1.00$^{(5)}$	0.92$^{(3)}$	1.33$^{(7)}$
MAR (S5)	15.80$^{(10)}$	7.04$^{(9)}$	1.22$^{(8)}$	**0.51**$^{(1)}$	0.55$^{(2)}$	0.99$^{(4)}$	0.71$^{(3)}$	1.15$^{(6)}$	1.00$^{(5)}$	1.20$^{(7)}$
S-MAR (S6)	16.07$^{(10)}$	6.84$^{(9)}$	1.25$^{(8)}$	**0.52**$^{(1)}$	0.56$^{(2)}$	0.91$^{(4)}$	0.73$^{(3)}$	1.15$^{(6)}$	1.02$^{(5)}$	1.22$^{(7)}$
MNAR (S7)	12.76	7.08	1.41	0.91	**0.82**	1.24	1.09	1.23	1.11	1.41
S-MNAR (S8)	12.96	7.25	1.47	0.90	**0.88**	1.19	1.08	1.24	1.08	1.39
					40% Missingness					
MCAR (S2)	16.70$^{(10)}$	1.16$^{(9)}$	0.77$^{(8)}$	**0.49**$^{(1)}$	0.59$^{(3)}$	0.58$^{(2)}$	0.68$^{(6)}$	0.60$^{(4)}$	0.65$^{(5)}$	0.69$^{(7)}$
MAR (S5)	4.76$^{(10)}$	0.23$^{(7)}$	0.22$^{(6)}$	0.20$^{(3)}$	0.24$^{(8)}$	0.22$^{(4)}$	0.28$^{(9)}$	**0.18**$^{(1)}$	0.22$^{(5)}$	0.20$^{(2)}$
S-MAR (S6)	5.12$^{(10)}$	0.25$^{(7)}$	0.25$^{(6)}$	0.23$^{(3)}$	0.24$^{(4)}$	0.26$^{(8)}$	0.34$^{(9)}$	**0.19**$^{(1)}$	0.25$^{(5)}$	0.23$^{(2)}$
MNAR (S7)	**0.88**	4.27	4.38	4.85	4.52	4.50	4.51	4.37	4.38	4.36
S-MNAR (S8)	**1.45**	4.16	4.00	4.50	4.13	4.18	4.20	3.97	4.03	4.03
					60% Missingness					
MCAR (S2)	16.36$^{(10)}$	1.72$^{(9)}$	1.19$^{(8)}$	**0.80**$^{(1)}$	0.94$^{(2)}$	1.12$^{(4)}$	1.25$^{(7)}$	1.17$^{(5)}$	1.12$^{(3)}$	1.18$^{(6)}$
MAR (S5)	0.38$^{(10)}$	0.18$^{(7)}$	0.12$^{(4)}$	0.14$^{(5)}$	0.16$^{(6)}$	0.19$^{(8)}$	0.21$^{(9)}$	0.11$^{(3)}$	0.08$^{(2)}$	**0.08**$^{(1)}$
S-MAR (S6)	0.60$^{(10)}$	0.15$^{(6)}$	0.12$^{(3)}$	0.15$^{(5)}$	0.16$^{(7)}$	0.21$^{(8)}$	0.23$^{(9)}$	0.14$^{(4)}$	0.10$^{(2)}$	**0.09**$^{(1)}$
MNAR (S7)	**5.60**	19.68	19.17	19.95	18.89	18.91	18.88	18.22	19.27	19.17
S-MNAR (S8)	**6.23**	19.00	17.75	18.38	17.36	17.46	17.47	16.80	17.80	17.70

Table 16: *Summary statistics of squared bias using adjusted weigths, analysed under the M1 simulation mechanism with second-order neighbors. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(0.5,0.008)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	20.58	6.22	1.34	**0.55**	0.61	0.81	0.74	/	/	/
					20% Missingness					
MCAR (S2)	21.71$^{(10)}$	8.01$^{(9)}$	1.71$^{(8)}$	**0.74**$^{(1)}$	0.82$^{(2)}$	1.01$^{(6)}$	0.93$^{(4)}$	1.00$^{(5)}$	0.92$^{(3)}$	1.33$^{(7)}$
MAR (S5)	15.80$^{(10)}$	7.04$^{(9)}$	1.22$^{(8)}$	**0.51**$^{(1)}$	0.55$^{(2)}$	0.99$^{(3)}$	0.71$^{(4)}$	1.15$^{(6)}$	1.00$^{(5)}$	1.20$^{(7)}$
S-MAR (S6)	16.07$^{(10)}$	6.84$^{(9)}$	1.25$^{(8)}$	**0.52**$^{(1)}$	0.56$^{(2)}$	0.91$^{(4)}$	0.73$^{(3)}$	1.15$^{(6)}$	1.02$^{(5)}$	1.22$^{(7)}$
MNAR (S7)	12.76	7.08	1.41	0.91	**0.82**	1.24	1.09	1.23	1.11	1.41
S-MNAR (S8)	12.96	7.25	1.47	0.90	**0.88**	1.19	1.08	1.24	1.08	1.39
					40% Missingness					
MCAR (S2)	23.05$^{(10)}$	10.49$^{(9)}$	2.17$^{(8)}$	**0.99**$^{(1)}$	1.14$^{(2)}$	1.23$^{(6)}$	1.20$^{(3)}$	1.21$^{(4)}$	1.23$^{(5)}$	1.64$^{(7)}$
MAR (S5)	11.74$^{(10)}$	8.98$^{(9)}$	1.12$^{(5)}$	**0.47**$^{(1)}$	0.47$^{(2)}$	1.16$^{(6)}$	0.65$^{(3)}$	1.23$^{(8)}$	1.09$^{(4)}$	1.21$^{(7)}$
S-MAR (S6)	11.64$^{(10)}$	8.24$^{(9)}$	1.18$^{(6)}$	**0.47**$^{(1)}$	0.55$^{(2)}$	1.10$^{(5)}$	0.71$^{(3)}$	1.26$^{(8)}$	1.10$^{(4)}$	1.20$^{(7)}$
MNAR (S7)	7.91	12.62	5.26	5.12	**4.75**	5.16	4.93	5.21	5.05	5.29
S-MNAR (S8)	8.53	12.69	5.02	4.78	**4.44**	4.83	4.66	4.79	4.69	4.96
					60% Missingness					
MCAR (S2)	26.09$^{(10)}$	15.47$^{(9)}$	2.71$^{(8)}$	**1.44**$^{(1)}$	1.46$^{(2)}$	1.77$^{(3)}$	1.81$^{(4)}$	1.93$^{(6)}$	1.86$^{(5)}$	2.27$^{(7)}$
MAR (S5)	11.44$^{(9)}$	13.01$^{(10)}$	1.26$^{(7)}$	**0.47**$^{(1)}$	0.58$^{(2)}$	0.98$^{(4)}$	0.65$^{(3)}$	1.29$^{(8)}$	1.08$^{(5)}$	1.18$^{(6)}$
S-MAR (S6)	11.92$^{(9)}$	12.91$^{(10)}$	1.09$^{(5)}$	0.48$^{(2)}$	**0.43**$^{(1)}$	1.01$^{(4)}$	0.67$^{(3)}$	1.36$^{(8)}$	1.11$^{(6)}$	1.25$^{(7)}$
MNAR (S7)	16.72	31.75	20.04	20.24	**19.11**	19.52	19.30	19.12	19.96	20.12
S-MNAR (S8)	18.11	31.63	18.73	18.80	**17.67**	18.23	18.07	17.77	18.65	18.75

Table 17: *Summary statistics of MSE using adjusted weigths, analysed under the M1 simulation mechanism with second-order neighbors. The prior distribution of the precision parameters σ_u^{-2} and σ_v^{-2} are assumed to follow a Gamma(0.5,0.008)-distribution. ($\times 10^3$)*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	**0.04**	0.04	1.03	0.34	1.88	0.32	0.33	/	/	/
					20% Missingness					
MCAR (S2)	**0.03**$^{(1)}$	0.04$^{(2)}$	2.06$^{(9)}$	0.51$^{(8)}$	2.84$^{(10)}$	0.48$^{(6)}$	0.48$^{(7)}$	0.48$^{(5)}$	0.48$^{(4)}$	0.46$^{(3)}$
MAR (S5)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.68$^{(9)}$	0.47$^{(8)}$	2.33$^{(10)}$	0.46$^{(5)}$	0.46$^{(7)}$	0.46$^{(6)}$	0.45$^{(4)}$	0.44$^{(3)}$
S-MAR (S6)	**0.04**$^{(1)}$	0.05$^{(2)}$	1.81$^{(9)}$	0.52$^{(8)}$	2.74$^{(10)}$	0.50$^{(6)}$	0.51$^{(7)}$	0.50$^{(5)}$	0.50$^{(4)}$	0.48$^{(3)}$
MNAR (S7)	**1.65**	1.69	3.16	2.03	4.24	2.00	1.99	2.27	1.98	1.96
S-MNAR (S8)	**1.59**	1.66	3.38	2.04	4.36	1.99	2.00	2.18	2.00	1.98
					40% Missingness					
MCAR (S2)	**0.05**$^{(1)}$	0.05$^{(2)}$	3.77$^{(9)}$	0.76$^{(4)}$	5.13$^{(10)}$	0.78$^{(6)}$	0.78$^{(7)}$	0.88$^{(8)}$	0.77$^{(5)}$	0.74$^{(3)}$
MAR (S5)	**0.04**$^{(1)}$	0.06$^{(2)}$	3.49$^{(9)}$	0.80$^{(5)}$	4.65$^{(10)}$	0.80$^{(6)}$	0.80$^{(7)}$	1.05$^{(8)}$	0.79$^{(4)}$	0.76$^{(3)}$
S-MAR (S6)	**0.06**$^{(1)}$	0.06$^{(2)}$	3.44$^{(9)}$	0.79$^{(4)}$	4.96$^{(10)}$	0.81$^{(8)}$	0.81$^{(7)}$	0.79$^{(5)}$	0.80$^{(6)}$	0.77$^{(3)}$
MNAR (S7)	8.55	**8.42**	11.64	9.16	12.46	9.26	9.17	10.13	9.12	9.12
S-MNAR (S8)	**8.36**	8.42	12.36	9.29	13.21	9.31	9.31	9.72	9.26	9.26
					60% Missingness					
MCAR (S2)	**0.08**$^{(1)}$	0.08$^{(2)}$	7.31$^{(9)}$	1.50$^{(3)}$	11.61$^{(10)}$	1.59$^{(6)}$	1.59$^{(7)}$	2.60$^{(8)}$	1.57$^{(5)}$	1.52$^{(4)}$
MAR (S5)	**0.08**$^{(1)}$	0.11$^{(2)}$	7.50$^{(9)}$	1.67$^{(7)}$	10.96$^{(10)}$	1.64$^{(6)}$	1.64$^{(5)}$	1.74$^{(8)}$	1.62$^{(4)}$	1.57$^{(3)}$
S-MAR (S6)	**0.07**$^{(1)}$	0.07$^{(2)}$	7.34$^{(9)}$	1.57$^{(4)}$	10.87$^{(10)}$	1.63$^{(8)}$	1.63$^{(7)}$	1.62$^{(6)}$	1.61$^{(5)}$	1.56$^{(3)}$
MNAR (S7)	28.40	**28.39**	39.47	31.80	38.81	31.97	31.94	32.74	31.82	31.85
S-MNAR (S8)	**27.86**	27.94	40.58	31.63	39.53	31.88	32.02	34.33	31.87	31.91

Table 18: *Summary statistics of squared bias using adjusted weigths, analysed under the M2 simulation mechanism with second-order neighbors. The prior distribution of the precision parameters* σ_u^{-2} *and* σ_v^{-2} *are assumed to follow a Gamma(0.5,0.008)-distribution.* ($\times 10^3$)

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Complete (S1)	2.46	3.63	6.72	3.49	9.34	**2.44**	2.45	/	/	/
					20% Missingness					
MCAR (S2)	3.16$^{(5)}$	4.65$^{(8)}$	9.29$^{(9)}$	4.44$^{(7)}$	9.83$^{(10)}$	**3.13**$^{(1)}$	3.15$^{(3)}$	3.16$^{(6)}$	3.16$^{(4)}$	3.14$^{(2)}$
MAR (S5)	2.98$^{(5)}$	4.27$^{(8)}$	8.41$^{(10)}$	4.08$^{(7)}$	8.30$^{(9)}$	**2.96**$^{(1)}$	2.98$^{(6)}$	2.96$^{(3)}$	2.97$^{(4)}$	2.96$^{(2)}$
S-MAR (S6)	**2.92**$^{(1)}$	4.15$^{(8)}$	8.56$^{(9)}$	4.03$^{(7)}$	10.10$^{(10)}$	2.95$^{(5)}$	2.96$^{(6)}$	2.94$^{(3)}$	2.95$^{(4)}$	2.94$^{(2)}$
MNAR (S7)	4.70	6.08	9.61	5.79	12.99	4.58	4.60	6.06	4.59	**4.58**
S-MNAR (S8)	4.59	6.13	9.98	5.87	12.36	**4.53**	4.56	5.56	4.56	4.56
					40% Missingness					
MCAR (S2)	4.15$^{(5)}$	5.96$^{(8)}$	12.78$^{(9)}$	5.57$^{(7)}$	14.43$^{(10)}$	4.07$^{(2)}$	4.09$^{(4)}$	4.46$^{(6)}$	4.08$^{(3)}$	**4.06**$^{(1)}$
MAR (S5)	**3.95**$^{(1)}$	5.38$^{(8)}$	12.36$^{(9)}$	5.16$^{(7)}$	13.13$^{(10)}$	3.97$^{(3)}$	4.01$^{(5)}$	5.08$^{(6)}$	3.98$^{(4)}$	3.96$^{(2)}$
S-MAR (S6)	3.86$^{(3)}$	5.20$^{(8)}$	12.42$^{(9)}$	4.98$^{(7)}$	14.93$^{(10)}$	3.86$^{(4)}$	3.88$^{(6)}$	3.85$^{(2)}$	3.87$^{(5)}$	**3.85**$^{(1)}$
MNAR (S7)	12.98	14.55	19.85	14.12	21.17	12.79	12.73	15.74	12.68	**12.68**
S-MNAR (S8)	12.60	14.55	20.84	14.26	21.37	12.72	12.75	14.15	**12.71**	12.72
					60% Missingness					
MCAR (S2)	6.40$^{(5)}$	8.64$^{(7)}$	19.64$^{(9)}$	7.90$^{(6)}$	27.49$^{(10)}$	6.20$^{(2)}$	6.22$^{(3)}$	9.40$^{(8)}$	6.22$^{(4)}$	**6.19**$^{(1)}$
MAR (S5)	6.25$^{(5)}$	8.03$^{(8)}$	19.43$^{(9)}$	7.60$^{(7)}$	25.37$^{(10)}$	6.19$^{(2)}$	6.22$^{(3)}$	6.74$^{(6)}$	6.22$^{(4)}$	**6.19**$^{(1)}$
S-MAR (S6)	6.26$^{(5)}$	7.83$^{(8)}$	19.36$^{(9)}$	7.46$^{(7)}$	25.55$^{(10)}$	6.24$^{(3)}$	6.26$^{(6)}$	6.23$^{(2)}$	6.25$^{(4)}$	**6.22**$^{(1)}$
MNAR (S7)	**36.43**	38.84	50.82	39.30	50.75	37.56	37.55	39.50	37.48	37.53
S-MNAR (S8)	**36.09**	38.77	51.30	39.08	51.58	37.39	37.56	42.49	37.45	37.50

Table 19: *Summary statistics of MSE using adjusted weigths, analysed under the M2 simulation mechanism with second-order neighbors. The prior distribution of the precision parameters* σ_u^{-2} *and* σ_v^{-2} *are assumed to follow a Gamma(0.5,0.008)-distribution.* ($\times 10^3$)

Health	Excellent	Good	Reasonable	Poor	Very Poor
Response	0	0	1	1	1

Table 20: *Dichotomization of the perceived health variable*

Description	Missingness probabilities q_{ik}^m														
Age Group	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
q_{ik}^m	0.23	0.18	0.09	0.07	0.08	0.06	0.07	0.08	0.08	0.07	0.08	0.07	0.10	0.18	0.28

Table 21: *Definition of the missingness probability weights for the simulation study of the HIS data set for each eligible age group.*

	UNW	HT	AN	PL	ES	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Bias^2	0.54	0.21	0.21	0.70	0.30	0.27	0.23	0.30	2.49	2.49
MSE	1.92	2.74	1.14	1.24	0.90	0.51	0.49	0.52	2.75	2.74

Table 22: *Summary statistics of squared bias and MSE using adjusted weigths for the simulation setting of the HIS data set. The prior distribution of the precision parameters* σ_u^{-2} *and* σ_v^{-2} *are assumed to follow a Gamma(0.5,0.008)-distribution.* ($\times 10^3$)

	MB1 (RW1)	MB1 (SP)	MB3 (RW1)	MB3 (SP)	MB3 (SP + OD)
Adjusted weights (W1)	10580.49	10576.65	10433.52	10214.34	/
Adjusted weights (W2)	10583.65	10583.82	10502.65	/	10388.13
Semi-adjusted weights	10586.64	10589.35	/	/	/

Table 23: *The DIC values of the hierarchical weight-smoothing estimators for the HIS dataset for the W1 and W2 approach in the calculation of the missingness weights.*

Figures

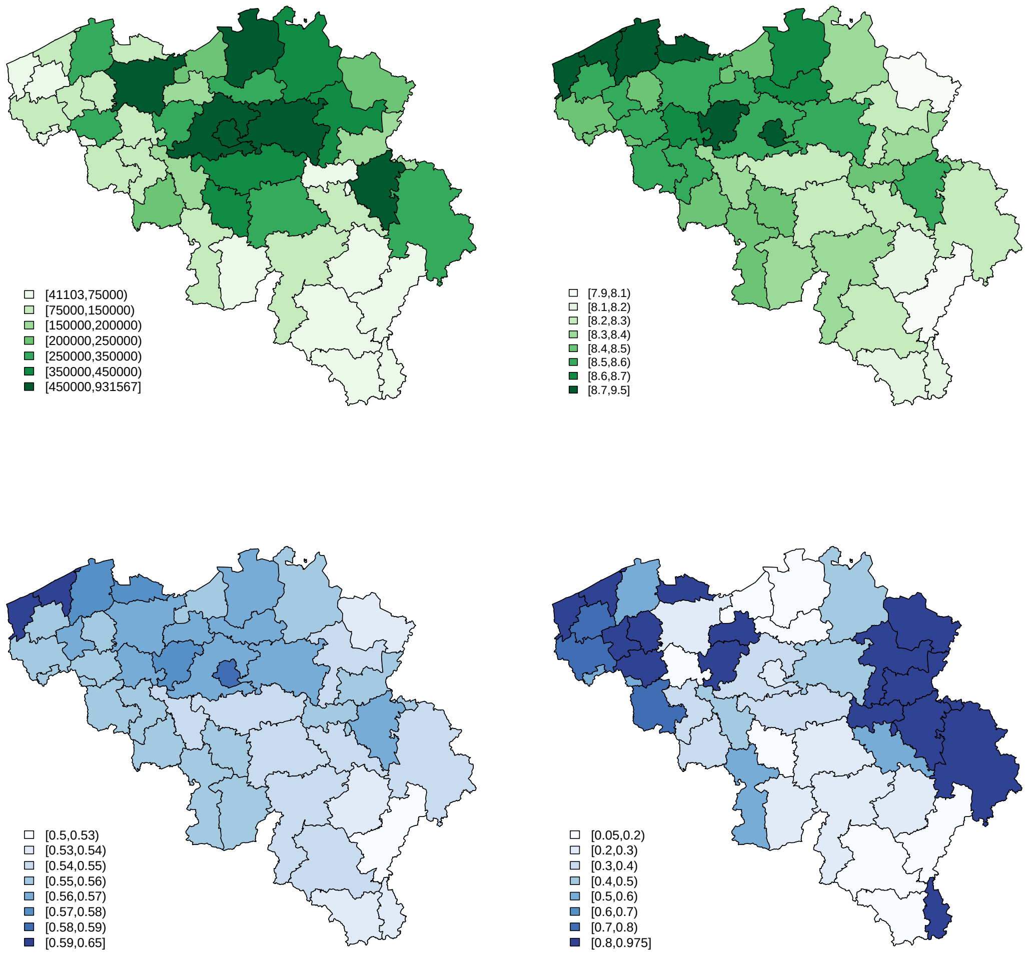

Figure 1: *The top panels show the population size (left panel) and average age category (right panel) per district. The lower panels correspond to the underlying prevalence models M1 (left) and M2 (right).*

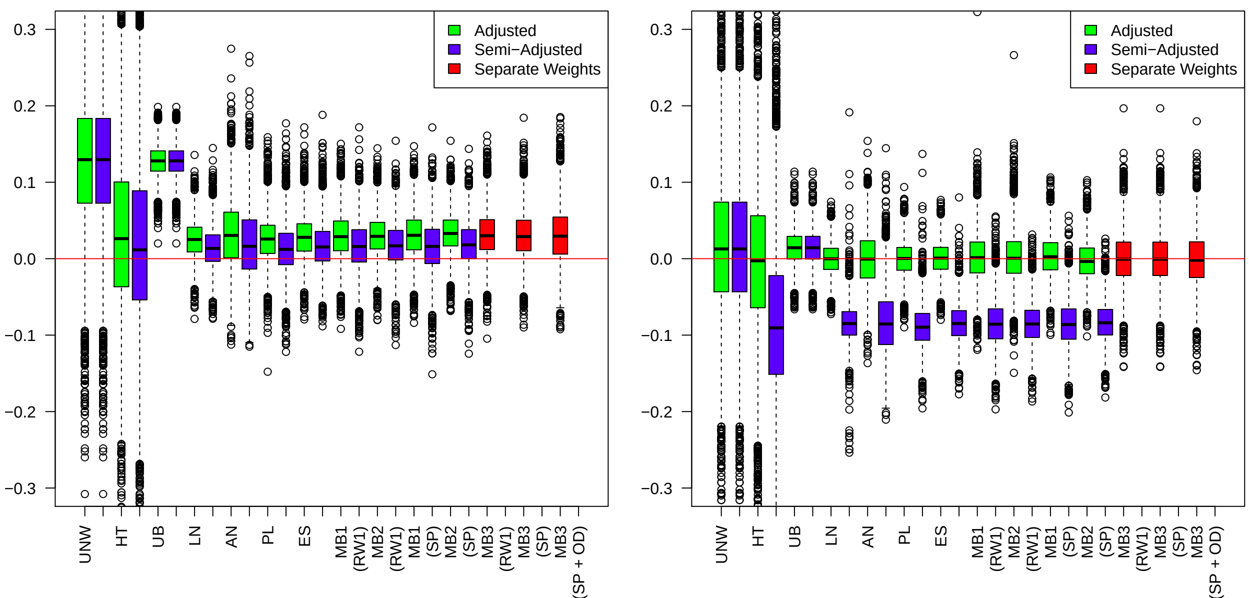

Figure 2: *The left panel depicts the box plots for the estimated bias under MCAR mechanism, while the right panel shows the box plots for those under the S-MAR assumption. Both were constructed under the M1 simulation mechanism, for 60% missingness.*

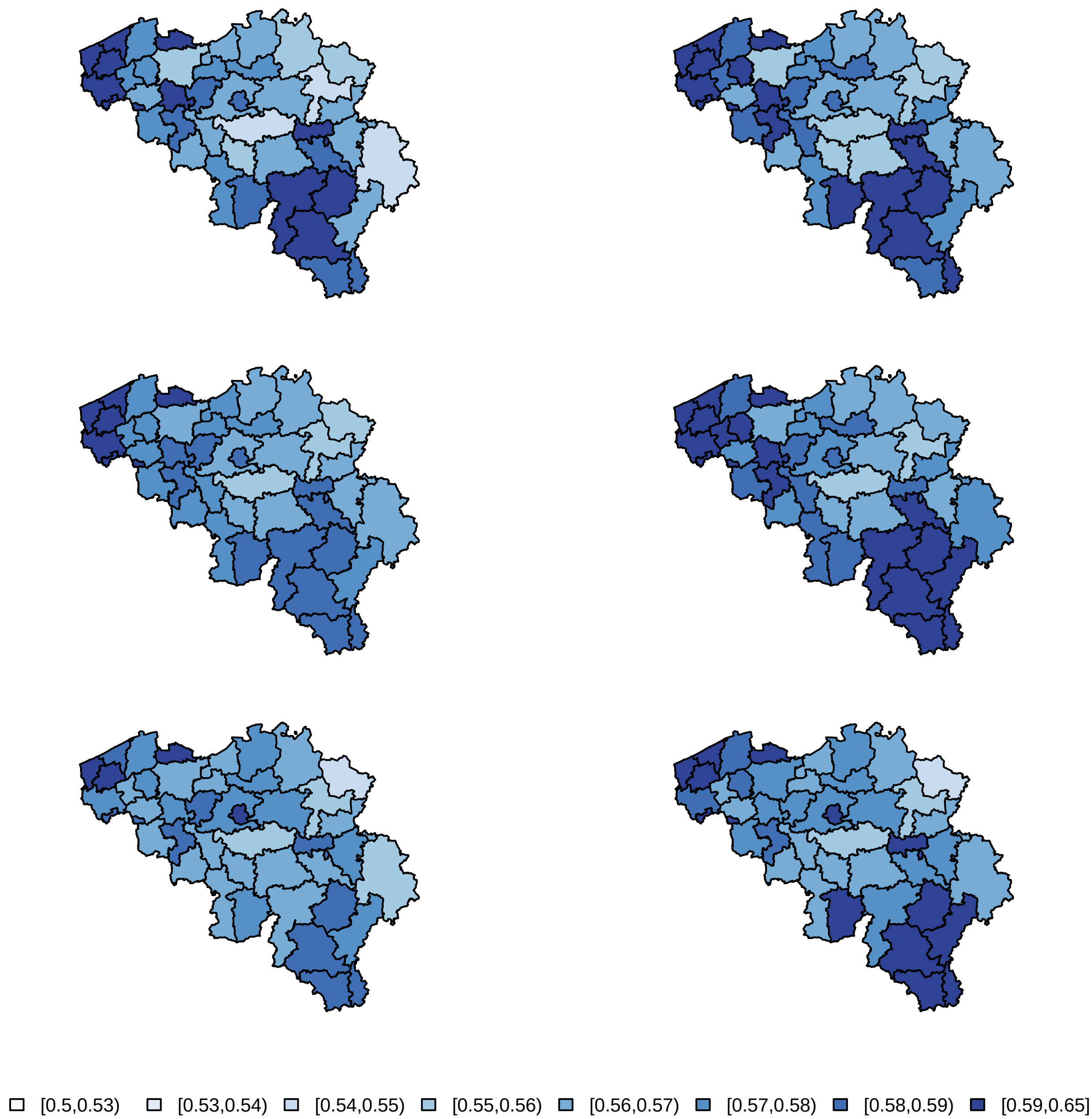

Figure 3: *Spatial maps displaying the estimated prevalence for the HT estimator (top row), AN estimator (middle row) and MB3 (SP+OD) estimator (bottom row), analysed under the M1 simulation mechanism and S2 (left column) and S6 (right column) missingness mechanism with 20% missingness.*

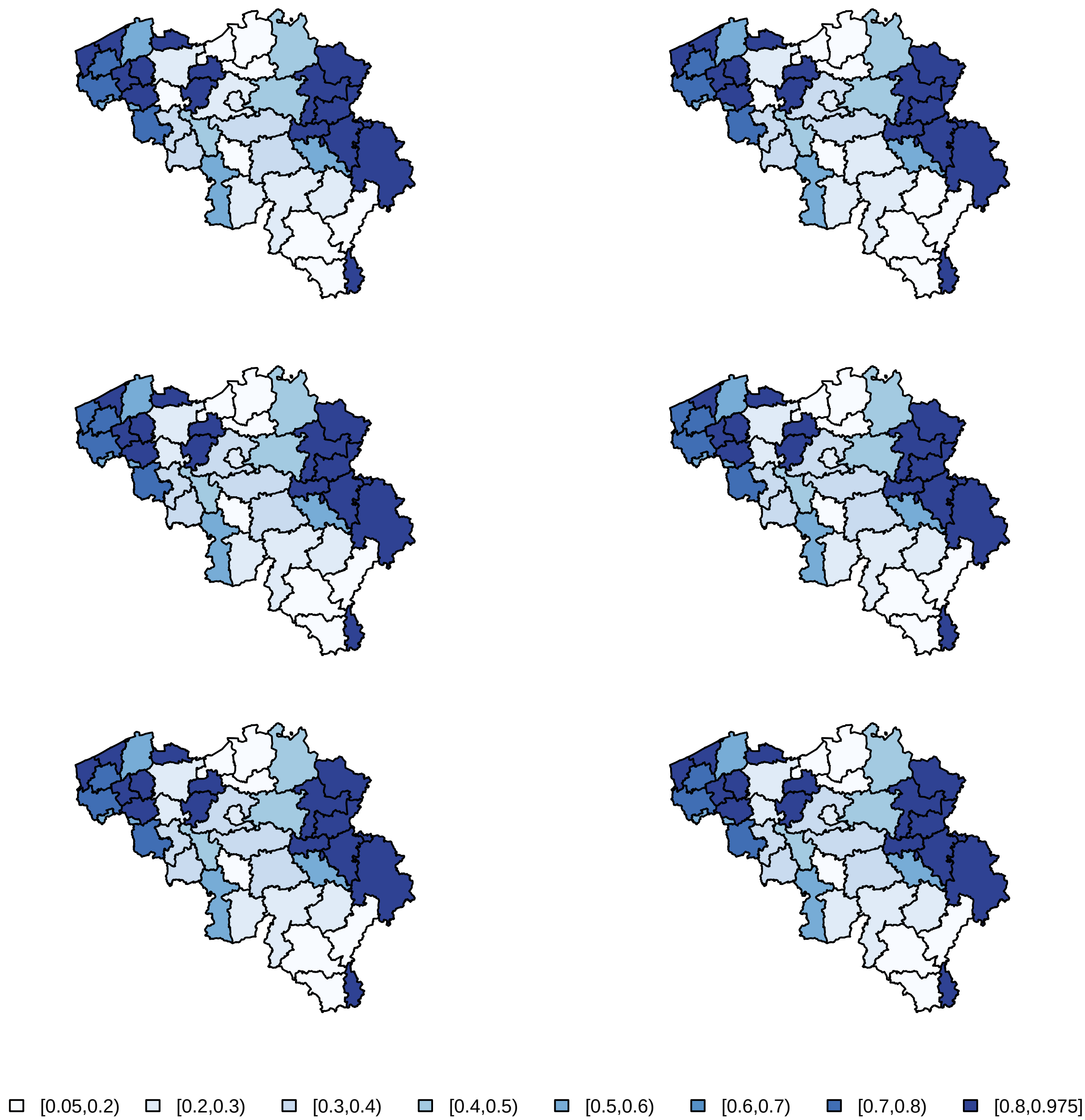

Figure 4: *Spatial maps displaying the estimated prevalence for the HT estimator (top row), AN estimator (middle row) and MB3 (SP+OD) estimator (bottom row), analysed under the M2 simulation mechanism and S2 (left column) and S6 (right column) missingness mechanism with 20% missingness.*

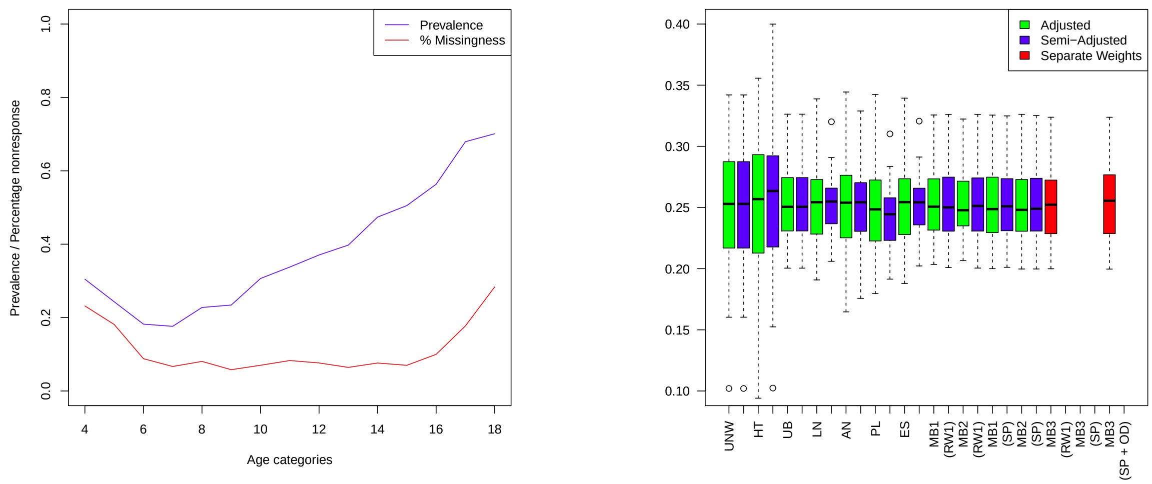

Figure 5: *Line plot for prevalence of perceived health and nonresponse percentages over five-year age intervals (left panel) and the box plots of the estimated prevalences for perceived health using the models with adjusted, semi-adjusted and separate weights (right panel).*

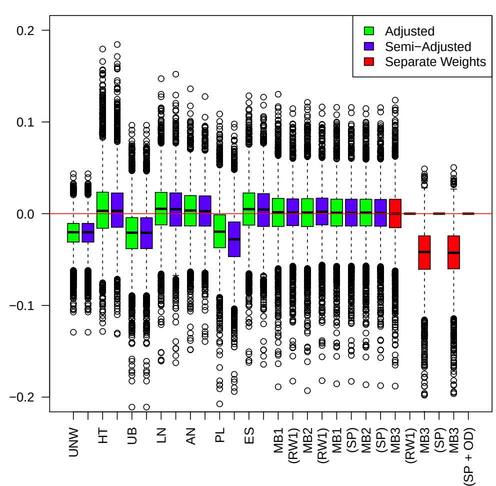

Figure 6: *Box plots of the simulated estimated prevalences for perceived health using the models with adjusted (W2), semi-adjusted and separate weights (right panel).*

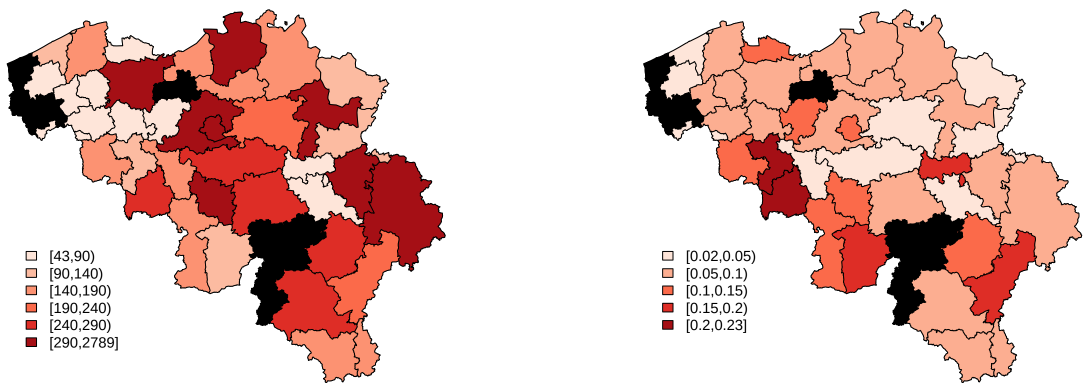

Figure 7: *Spatial maps of the sample sizes for the Belgian districts (left) and the amount of missingness for the perceived health variable in HIS data set (right). Unsampled municipalities are highlighted in black.*

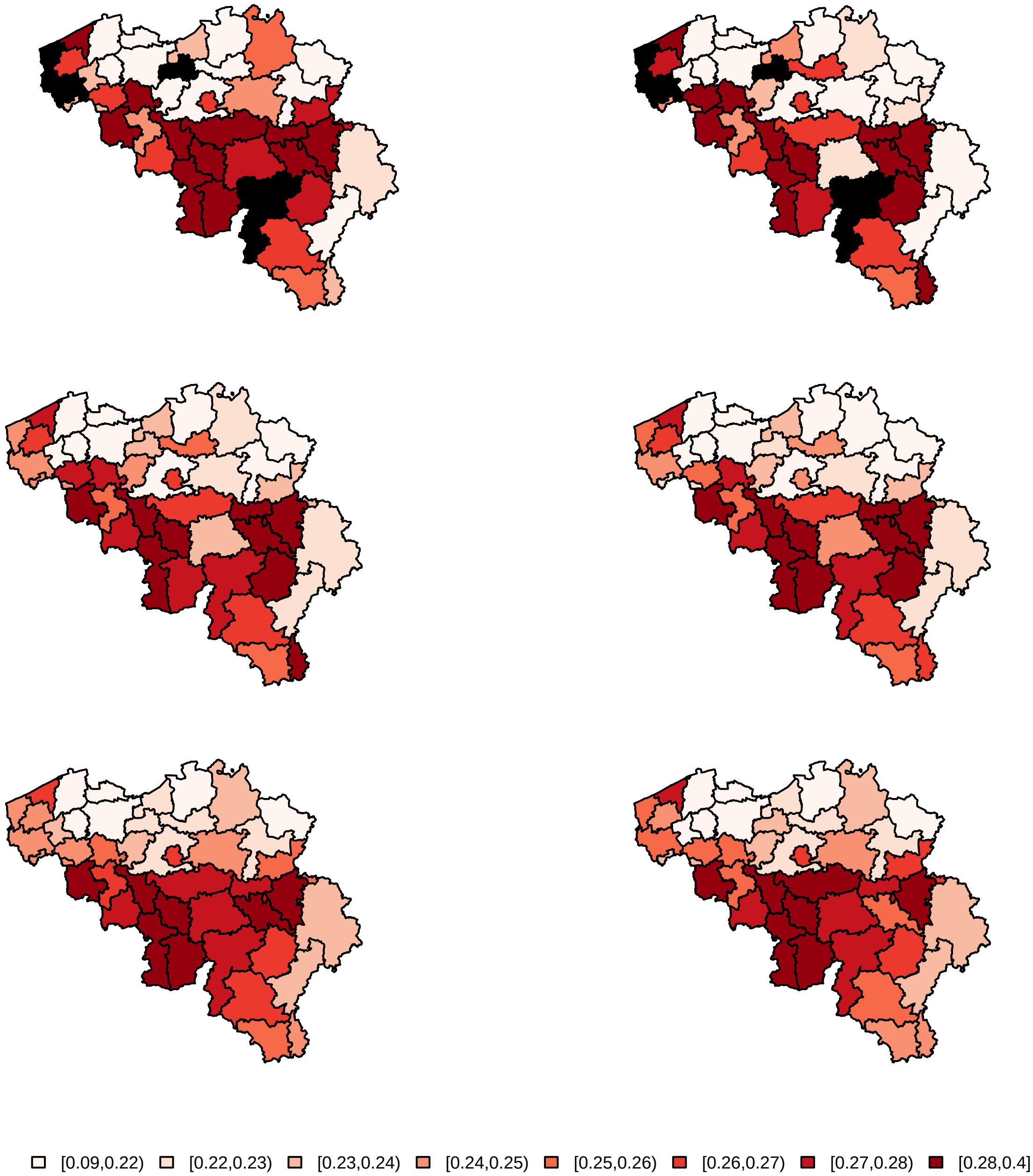

Figure 8: *Spatial distribution of UNW estimates (top left), HT estimates (top right), AN estimates, (middle left), PL estimates (middle right), MB1 (SP) estimates (bottom left) and MB3 (RW1) estimates (Bottom right) using adjusted weights. Unsampled municipalities are highlighted in black for the UNW and HT estimates.*