Handling Missing Data in Matched Case-Control Studies using Multiple Imputation

Shaun R. Seaman[1]*, Ruth H. Keogh[2]

[1] MRC Biostatistics Unit, Cambridge, CB2 0SR, UK

[2] London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

**email:* shaun.seaman@mrc-bsu.cam.ac.uk

Summary:

Analysis of matched case-control studies is often complicated by missing data on covariates. Analysis can be restricted to individuals with complete data, but this is inefficient and may be biased. Multiple imputation (MI) is an efficient and flexible alternative. We describe two MI approaches. The first uses a model for the data on an individual and includes matching variables; the second uses a model for the data on a whole matched set and avoids the need to model the matching variables. Within each approach, we consider three methods: full-conditional specification (FCS), joint model MI using a normal model, and joint model MI using a latent normal model. We show that FCS MI is asymptotically equivalent to joint model MI using a restricted general location model that is compatible with the conditional logistic regression analysis model. The normal and latent normal imputation models are not compatible with this analysis model. All methods allow for multiple partially-observed covariates, non-monotone missingness and multiple controls per case. They can be easily applied in standard statistical software and valid variance estimates obtained using Rubin's Rules. We compare the methods in a simulation study. The approach of including the matching variables is most efficient. Within each approach, the FCS MI method generally yields the least biased odds ratio estimates, but normal or latent normal joint model MI is sometimes more efficient. All methods have good confidence interval coverage. Data on colorectal cancer and fibre intake from the EPIC-Norfolk study are used to illustrate the methods, in particular showing how efficiency is gained relative to just using individuals with complete data.

Key words: chained equations, compatibility, MICE, multilevel MI, restricted general location model

1. Introduction

Case-control studies are used to investigate associations between disease and putative risk factors. Confounding of observed associations can be handled at the design stage by matching cases and controls on confounders, at the analysis stage by adjusting for confounders using a regression model, or by a combination of these. In matched case-control studies each case is individually matched with one or more controls on a subset of confounders and the (usual) analysis uses conditional logistic regression (CLR) to control for the remaining confounders.

Often, the analysis is complicated by missing data on covariates (i.e. exposures and remaining confounders). A common solution is to restrict analysis to individuals with complete data. Although appealing for its simplicity, this 'complete-case analysis' ('case' here means any individual, rather than an individual with disease) is inefficient and may be biased. In particular, where exclusion of a case or control leaves a matched set in which remaining members are either all cases or all controls, the whole set ceases to contribute information to the CLR estimating equations.

To improve efficiency and reduce bias, several alternatives have been proposed. Lipsitz et al. (1998) allow for one partially observed covariate. They assume data are missing at random (MAR) and fit a missingness model, i.e. a model for the probability that an individual is a complete case. Functions of the fitted probabilities are then used as offsets in CLR. This consistently estimates odds ratios (ORs) when the missingness model is correctly specified, but is inefficient as it only uses data on complete cases. Paik and Sacco (2000) also allow for just one partially observed covariate and assume MAR. They assume a model for the distribution of the partially observed covariate given the other covariates, matching variables and binary disease status. When this covariate model is correctly specified, consistent estimation of the ORs can be achieved by CLR after imputing the missing covariate as its fitted value when the disease status variable is set to 0.5. Rathouz (2003) notes that this

method implicitly assumes missingness does not depend on disease status, and generalises it to allow for such dependence, as well as for multiple missing covariates. His method assumes the partially observed covariates are all observed or all missing on each individual. Sinha and Wang (2009) take a similar approach, but instead of a parametric covariate model, kernel density estimation is used for those functions in the estimating equations that depend on the distribution of the partially observed covariate. They find their OR estimator is less biased than that of Paik and Sacco (2000) when the latter's covariate model is misspecified. A drawback is that categorical variables are handled by stratifying individuals on these variables and performing kernel density estimation separately in each stratum, which limits the feasible number of categorical variables (and categories). Paik (2004) extends Paik and Sacco's (2000) method to allow for data missing not at random (MNAR).

The forementioned methods all reduce to standard CLR when there are no missing data: the assumed missingness or covariate model then becomes irrelevant. Other methods for missing data derive information from an assumed covariate model even when data are complete. These methods may be more efficient but at the cost of possible bias when the covariate model is misspecified. Satten and Carroll (2000) propose such a method. This allows for multiple partially observed covariates, but assumes these are all observed or all missing on each individual. Ahn et al. (2011) generalise it to allow for MNAR and multiple disease states. Rathouz et al. (2002) elaborate Lipsitz et al.'s (1998) method to use a covariate model and so gain efficiency. The resulting estimator is doubly robust but difficult to implement. They also propose a more practical approximation which, though not doubly robust, still gains efficiency. Liu et al. (2013) use empirical likelihood to develop a semiparametric-efficient competitor to Rathouz et al.'s (2002) estimator. Gebregziabher and DeSantis (2010) assume all covariates are categorical and carry out multiple imputation (MI) using a latent-class model. A drawback is that imputation of a individual's missing value makes no use of data

on matching variables, covariate values of other individuals in the same matched set, or disease status, which may cause bias (Moons et al., 2006) and inefficiency.

The methods described so far assume the distribution of the partially observed covariate(s) given fully observed covariates, disease status and matching variables can be modelled parametrically. Sometimes this is not feasible. For example, if cases are matched with controls from the same family, from the same postcode area or from the set of patients attending the same general practice, it could be difficult to model parametrically the matching via explicit matching variables, while the alternative of allowing a separate nuisance parameter for each matched set may cause problems with model fitting and induce bias and even inconsistency of estimators. Even when matching could, in principle, be modelled parametrically, this is only possible if the analyst has data on matching variables, which is not always so, and some analysts may prefer to avoid modelling effects of matching variables, since CLR makes no assumptions about the association between disease and matching variables. One solution, adopted by Sinha et al. (2005), is to allow each matched set to have its own parameter in the covariate model but treat these as random effects. They assume a single partially observed covariate and that the random effects are generated by a Dirichlet process. They fit their Bayesian model using a Hastings-Metropolis algorithm with specially written computer code.

Though useful, these methods have limitations. Many assume only one partially observed covariate or that partially observed covariates are collectively observed or missing on each individual. Many require bespoke computer code. Most require parametric modelling of matching variables. In this article we advocate the use of MI, proposing and comparing six MI methods suitable for matched case-control data that can be easily implemented in commonly used statistical packages. MI has several advantages. First, it is increasingly being used to handle missing data and many researchers are familiar with the technique. Second, MI software is readily available and easy to use. Third, MI allows for multiple partially

observed covariates without needing them to be collectively observed or missing. Fourth, MI can easily incorporate information on variables that are not included in the CLR model but are predictive of missing covariates in that model. This can increase efficiency and can also reduce bias when these extra variables are required to make the MAR assumption more plausible. Fifth, we propose both methods that parametrically model matching variables and methods in which this is not required. Arguably, a sixth advantage is that, unlike some of the methods proposed earlier, MI reduces to standard CLR when there are no missing data. Although this means MI does not offer the potential efficiency gain associated with methods that make use of a covariate model even when data are complete, it should make it more robust to misspecification of that model.

We illustrate the use of MI for matched case-control data on a study of association between fibre intake and colerectal cancer nested within the European Prospective Investigation of Cancer (EPIC) Norfolk cohort. This is one of the studies in the UK Dietary Cohort Consortium, which combines case-control studies nested within several cohorts. Results from this study have been described elsewhere (Dahm et al., 2010). Cases were individuals in the EPIC Norfolk cohort diagnosed with colorectal cancer between recruitment to the cohort (1993–1998) and the end of 2006. Seven-day diet diaries were completed by participants shortly after recruitment to the underlying cohort and stored for later use. The diet diaries were processed for individuals selected for the case-control sample to obtain measures of average daily intake of foods and nutrients (Dahm et al., 2010). There were 318 colorectal cancer cases and each was matched with 4 controls on sex, age within 3 years, and date of diary completion within 3 months. Controls had to be alive and have not been diagnosed with colorectal cancer at the end of 2006. In the original analysis, the association between fibre intake and colorectal cancer was adjusted for several potential confounders using CLR : smoking status (3 categories), education (4 categories), social class (6 categories) and physical

activity level (4 categories), and height, weight, exact age, alcohol intake, folate intake, intake of energy from fat and intake of energy from non-fat (all continuous). We wished also to adjust for aspirin use (2 categories). Many other studies have adjusted for aspirin use (Aune et al., 2011), which is known to be associated with reduced risk of colorectal cancer (Asano and McLeod, 2004; National Cancer Institute, 2014). It was omitted from the original analysis (Dahm et al., 2010) because it was not measured in some of the contributing studies. Of the 1590 individuals in the study, 328 (78 cases and 250 controls) were missing one or more adjustment variables, most commonly aspirin use or social class; the main exposure, fibre intake, and the matching variables were fully observed. A complete-case analysis uses only 240 (75%) matched sets and 1012 (64%) individuals.

The article is structured as follows. Section 2 discusses CLR with complete data. Section 3 describes MI in general. For matched case-control studies, Section 4 proposes three MI methods that parametrically model the matching variables, and Section 5 three analogous methods that avoid this. Section 6 contains a simulation study comparing the methods. Section 7 describes their application to the EPIC study. We end with a discussion in Section 8.

2. Analysis of Matched Case-Control Studies with Complete Data

For each individual in the population, let $D = 1$ if he/she has disease and $D = 0$ otherwise. So, $D = 1$ for cases and $D = 0$ for controls. Let $\boldsymbol{S}$ denote the variables used to match controls with cases. Let $\boldsymbol{X}^{\text{cat}}$ and $\boldsymbol{X}^{\text{con}}$ denote categorical and continuous covariates, respectively. A categorical variable with $m > 2$ levels is coded as $m - 1$ dummy variables. Assume

$$P(D = 1 \mid \boldsymbol{X}^{\text{cat}}, \boldsymbol{X}^{\text{con}}, \boldsymbol{S}) = \frac{\exp\{\boldsymbol{\beta}_{\text{cat}}^{\top}\boldsymbol{X}^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^{\top}\boldsymbol{X}^{\text{con}} + q(\boldsymbol{S})\}}{1 + \exp\{\boldsymbol{\beta}_{\text{cat}}^{\top}\boldsymbol{X}^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^{\top}\boldsymbol{X}^{\text{con}} + q(\boldsymbol{S})\}}, \tag{1}$$

where $q(\boldsymbol{S}) = \text{logit } P(D = 1 \mid \boldsymbol{X}^{\text{cat}} = \boldsymbol{0}, \boldsymbol{X}^{\text{con}} = \boldsymbol{0}, \boldsymbol{S})$. Let M denote the number of controls matched with each case. We use subscript j ($j = 1, \ldots, M+1$) to index individual within set and assume cases and controls have been ordered so that $D_1 = 1$ and $D_2 = \ldots = D_{M+1} = 0$.

In ordinary logistic regression the log ORs $\boldsymbol{\beta}_{\text{cat}}$ and $\boldsymbol{\beta}_{\text{con}}$ are estimated by maximising the likelihood based on expression (1) and the data on the sampled individuals. This requires that either $q(\boldsymbol{S})$ is modelled or a separate baseline parameter is included for each matched set. The former corresponds to breaking the matching and adjusting for $\boldsymbol{S}$, which there is often a reluctance to do, because it requires a functional form to be specified for the effect of matching variables on disease risk. The alternative of including a baseline parameter for each set yields inconsistent maximum likelihood estimates (Breslow and Day, 1980). For this reason, CLR is often used instead. CLR includes a baseline parameter for each set, but then eliminates these from the likelihood by conditioning on the number of cases and controls in each set. Let $G(\boldsymbol{x}_1^{\text{cat}}, \boldsymbol{x}_1^{\text{con}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}}, \boldsymbol{x}_{M+1}^{\text{con}})$ denote the conditional probability that $(\boldsymbol{X}_1^{\text{cat}}, \boldsymbol{X}_1^{\text{con}}) = (\boldsymbol{x}_1^{\text{cat}}, \boldsymbol{x}_1^{\text{con}})$ given that $(\boldsymbol{X}_1^{\text{cat}}, \boldsymbol{X}_1^{\text{con}}) = (\boldsymbol{x}_1^{\text{cat}*}, \boldsymbol{x}_1^{\text{con}*}), \ldots, (\boldsymbol{X}_{M+1}^{\text{cat}}, \boldsymbol{X}_{M+1}^{\text{con}}) = (\boldsymbol{x}_{M+1}^{\text{cat}*}, \boldsymbol{x}_{M+1}^{\text{con}*})$ for some permutation $(\boldsymbol{x}_1^{\text{cat}*}, \boldsymbol{x}_1^{\text{con}*}), \ldots, (\boldsymbol{x}_{M+1}^{\text{cat}*}, \boldsymbol{x}_{M+1}^{\text{con}*})$ of $(\boldsymbol{x}_1^{\text{cat}}, \boldsymbol{x}_1^{\text{con}}), \ldots, (\boldsymbol{x}_{M+1}^{\text{cat}}, \boldsymbol{x}_{M+1}^{\text{con}})$ and given that $D_1 = 1$ and $D_2 = \ldots = D_{M+1} = 0$ and $\boldsymbol{S}_1 = \ldots = \boldsymbol{S}_{M+1}$. Equation (1) implies

$$G(\boldsymbol{x}_1^{\text{cat}}, \boldsymbol{x}_1^{\text{con}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}}, \boldsymbol{x}_{M+1}^{\text{cat}}) = \frac{\exp(\boldsymbol{\beta}_{\text{cat}}^{\top}\boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^{\top}\boldsymbol{x}_1^{\text{con}})}{\sum_{j=1}^{M+1}\exp(\boldsymbol{\beta}_{\text{cat}}^{\top}\boldsymbol{x}_j^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^{\top}\boldsymbol{x}_j^{\text{con}})}, \tag{2}$$

and vice versa (Web Appendix A). CLR finds the values of $\boldsymbol{\beta}_{\text{cat}}$ and $\boldsymbol{\beta}_{\text{con}}$ that maximise the product of expression (2) over the matched sets; these consistently estimate the log ORs .

3. Joint model MI and full-conditional specification (FCS) MI

We briefly review the most commonly used forms of MI: joint model MI and FCS MI (Web Appendix B has more detail). In joint model MI, a Bayesian model with non-informative priors is specified for the distribution of the partially observed variables given fully observed variables. This 'imputation model' is fitted to the observed data, and values for missing variables then sampled from their joint posterior predictive distribution. The model of interest ('analysis model') is fitted to each resulting complete (or 'imputed') dataset separately, and the parameter and variance estimates obtained are combined using simple equations

called Rubin's Rules. When the imputation model is correctly specified and is *compatible with the analysis model*, i.e. there exists a model for the joint distribution of all the variables that implies the analysis and imputation models as submodels, and data are MAR, joint model MI gives consistent parameter and variance estimates for the analysis model. Thus, compatibility, if possible, is desirable. The first of the methods described in each of Sections 4 and 5 are based on imputation models that are compatible with the CLR analysis model.

Instead of requiring a joint model for the partially observed variables, FCS MI involves specifying a model for the conditional distribution of each partially observed variable given all other variables. The FCS algorithm cycles through these models, sampling missing values for the dependent variable in the current model given the observed and most recently sampled values of all the other variables, until convergence is achieved. This may be easier than specifying and fitting a joint model. In special cases FCS corresponds to joint model MI (Hughes et al., 2014). Otherwise, FCS is less theoretically justified, but there is much evidence that it works well in terms of approximate unbiasedness of parameter and variance estimates and coverage of confidence intervals (van Buuren, 2012; Hughes et al., 2014; Lee and Carlin, 2010). An important theoretical result was given by Liu et al. (2014). They defined the set of conditional models to be *compatible with a joint model* if, for each conditional model and every possible set of parameter values for that model, there exists a set of parameter values for the joint model such that the conditional and joint models imply the same distribution for the dependent variable of that conditional model. They showed that when this compatibility holds, the distribution of the data imputed by FCS MI converges, as sample size tends to infinity, to the posterior predictive distribution of the missing data under that joint model. Hence, FCS MI is asymptotically equivalent to joint model MI in this case. The first of the MI methods in each of Sections 4 and 5 use this asymptotic result.

4. MI using matching variables

Let $\boldsymbol{R}$ denote the missingness pattern in $(\boldsymbol{X}^{\text{cat}\top}, \boldsymbol{X}^{\text{con}\top})^\top$. Assume D and $\boldsymbol{S}$ are fully observed and the data are MAR. In this section, we propose multiply imputing missing $(\boldsymbol{X}^{\text{cat}\top}, \boldsymbol{X}^{\text{con}\top})^\top$ from its conditional distribution given $\boldsymbol{S}$ and D. We call this 'MI using matching variables'. It is analogous to breaking the matching and adjusting for the matching variables. However, matching is broken only to impute missing data; matching is then restored and the imputed data analyzed using CLR. Most methods reviewed in Section 1 effectively break the matching for the individuals with missing data. In Section 5 we describe an alternative ('MI using matched set'), which imputes without breaking the matching. We now propose three ways of modelling the distribution of $(\boldsymbol{X}^{\text{cat}\top}, \boldsymbol{X}^{\text{con}\top})^\top$ given $\boldsymbol{S}$ and D.

The first model for $(\boldsymbol{X}^{\text{cat}\top}, \boldsymbol{X}^{\text{con}\top})^\top$ given $\boldsymbol{S}$ and D is a restricted general location model (Schafer, 1997). This has a log-linear model for $\boldsymbol{X}^{\text{cat}}$ and normal model for $\boldsymbol{X}^{\text{con}}$ given $\boldsymbol{X}^{\text{cat}}$:

$$P(\boldsymbol{X}^{\text{cat}} = \boldsymbol{x}^{\text{cat}} \mid \boldsymbol{S}, D) = \frac{\exp\{a(\boldsymbol{x}^{\text{cat}}, \boldsymbol{S};\ \boldsymbol{\zeta}) + D\boldsymbol{\lambda}^\top \boldsymbol{x}^{\text{cat}}\}}{\sum_{\boldsymbol{x}^{\text{cat}\prime}} \exp\{a(\boldsymbol{x}^{\text{cat}\prime}, \boldsymbol{S};\ \boldsymbol{\zeta}) + D\boldsymbol{\lambda}^\top \boldsymbol{x}^{\text{cat}\prime}\}} \tag{3}$$

$$\boldsymbol{X}^{\text{con}} \mid \boldsymbol{X}^{\text{cat}}, \boldsymbol{S}, D \sim N(\boldsymbol{\alpha} + \boldsymbol{\phi}D + \boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} + \boldsymbol{\delta}\boldsymbol{S},\ \boldsymbol{\Sigma}) \tag{4}$$

where $a(\boldsymbol{x}^{\text{cat}}, \boldsymbol{S};\ \boldsymbol{\zeta})$ includes a main effect for $\boldsymbol{X}^{\text{cat}}$ and all pairwise interactions between $\boldsymbol{X}^{\text{cat}}$ and $\boldsymbol{S}$ and between pairs of elements of $\boldsymbol{X}^{\text{cat}}$. Vectors $\boldsymbol{\lambda}$, $\boldsymbol{\zeta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ and matrices $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}$ are unknown parameters. In Web Appendix C, we prove that (3)–(4) imply that equation (2) holds with $\boldsymbol{\beta}_{\text{cat}} = \boldsymbol{\lambda} - \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\phi}$ and $\boldsymbol{\beta}_{\text{con}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}$. Hence, this model is compatible with the CLR analysis model.

Bayesian modelling software, such as WinBUGS (Lunn et al., 2000), can be used to impute missing $(\boldsymbol{X}^{\text{cat}\top}, \boldsymbol{X}^{\text{con}\top})^\top$ from its posterior predictive distribution implied by joint model (3)–(4). However, such software requires specialist programming skills. Instead we propose using FCS MI with a set of conditional models that is compatible with this joint model, and hence is asymptotically equivalent to joint model MI. FCS MI is widely available in general-purpose

statistical packages, e.g. Stata, R and SAS. In Web Appendix D, we show that a compatible conditional model for a partially observed continuous covariate (an element of $\boldsymbol{X}^{\text{con}}$) is a linear regression of this covariate on $\boldsymbol{S}$, D, $\boldsymbol{X}^{\text{cat}}$ and the remaining elements of $\boldsymbol{X}^{\text{con}}$. Likewise, a compatible conditional model for one of the partially observed categorical covariates making up $\boldsymbol{X}^{\text{cat}}$ is a multinomial logistic regression of this categorical covariate on $\boldsymbol{S}$, D, $\boldsymbol{X}^{\text{con}}$ and those elements of $\boldsymbol{X}^{\text{cat}}$ that are not dummy indicators for this categorical covariate. Conveniently, these conditional models are the default options in many MI packages.

Although asymptotically equivalent, in finite samples this FCS MI method may be inefficient compared to joint model MI, because it estimates the parameters of the conditional model for $\boldsymbol{X}^{\text{cat}}$ using only part of the available data on $\boldsymbol{X}^{\text{con}}$ (Hughes et al., 2014). Our second proposed model for $(\boldsymbol{X}^{\text{cat}\top}, \boldsymbol{X}^{\text{con}\top})^{\top}$ given $\boldsymbol{S}$ and D is a latent normal model (Carpenter and Kenward, 2013). This is not compatible with the CLR analysis model, but it has the advantage that it can be used for joint model MI without needing specialist Bayesian software. For simplicity, suppose that all the categorical covariates are binary (see Carpenter and Kenward (2013) for general case). The latent normal model is

$$(\boldsymbol{X}^{\text{con}\top}, \boldsymbol{W}^{\text{cat}\top})^{\top} \mid \boldsymbol{S}, D \sim N(\boldsymbol{\alpha}_{\text{LN}} + \boldsymbol{\phi}_{\text{LN}} D + \boldsymbol{\delta}_{\text{LN}} \boldsymbol{S},\ \boldsymbol{\Sigma}_{\text{LN}}) \tag{5}$$

where $\boldsymbol{W}^{\text{cat}}$ is a vector of latent variables (each with unit variance), one for each element of $\boldsymbol{X}^{\text{cat}}$, and such that an element of $\boldsymbol{X}^{\text{cat}}$ equals 1 if its corresponding element of $\boldsymbol{W}^{\text{cat}}$ is positive and 0 otherwise. $\boldsymbol{\alpha}_{\text{LN}}$, $\boldsymbol{\phi}_{\text{LN}}$, $\boldsymbol{\delta}_{\text{LN}}$ and $\boldsymbol{\Sigma}_{\text{LN}}$ are unknown parameters. Joint model MI using (5) can be done using the software REALCOM-MI or the jomo package in R. The realcomImpute program provides an interface between Stata and REALCOM-MI.

When all partially observed covariates are continuous, our FCS MI method and joint model MI using (5) both reduce to joint model MI using the normal model (4). Use of this normal model for MI even when some partially observed variables are categorical was originally promoted by Schafer (1997) and has become common. Although the model is obviously

misspecified, this method has been found to work well in many situations and software is widely available, e.g. mi mvn impute in Stata and norm in R. Thus, our third proposed model for $(\boldsymbol{X}^{\text{cat}\top}, \boldsymbol{X}^{\text{con}\top})^\top$ given $\boldsymbol{S}$ and D is expression (5) with $\boldsymbol{W}^{\text{cat}}$ replaced by $\boldsymbol{X}^{\text{cat}}$. Following Bernaards et al. (2007), we use 'adaptive rounding' after imputation to handle non-integer imputed values of $\boldsymbol{X}^{\text{cat}}$.

5. MI using matched set

Now we propose three models for $\boldsymbol{X}^{\text{set}} = (\boldsymbol{X}_1^{\text{cat}\top}, \boldsymbol{X}_1^{\text{con}\top}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}\top}, \boldsymbol{X}_{M+1}^{\text{con}\top})^\top$ unconditional on $\boldsymbol{S}$, thus allowing imputation without using the matching variables. These are analogous to the models in Section 4 but involve a matched-set-specific random effect, $\boldsymbol{u}$.

The first is a restricted general location model. Assume that for each matched set,

$$P(\boldsymbol{X}_1^{\text{cat}} = \boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}} = \boldsymbol{x}_{M+1}^{\text{cat}} \mid D_1 = 1, D_2 = \ldots = D_{M+1} = 0) = \frac{\exp\{b(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}};\ \boldsymbol{\nu}) + \boldsymbol{\tau}^\top \boldsymbol{x}_1^{\text{cat}}\}}{\sum_{\boldsymbol{x}_1^{\text{cat}\prime}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}\prime}} \exp\{b(\boldsymbol{x}_1^{\text{cat}\prime}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}\prime};\ \boldsymbol{\nu}) + \boldsymbol{\tau}^\top \boldsymbol{x}_1^{\text{cat}\prime}\}} \tag{6}$$

$$\text{with } b(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}}) = \sum_{j=1}^{M+1} b_1(\boldsymbol{x}_j^{\text{cat}};\ \boldsymbol{\nu}) + \sum_{j=1}^{M} \sum_{k=j+1}^{M+1} b_2(\boldsymbol{x}_j^{\text{cat}}, \boldsymbol{x}_k^{\text{cat}}; \boldsymbol{\nu}) \tag{7}$$

where $b_1(\boldsymbol{x}_j^{\text{cat}};\ \boldsymbol{\nu})$ includes a main effect of each element of $\boldsymbol{X}_j^{\text{cat}}$ and an interaction between each pair of these elements, and $b_2(\boldsymbol{x}_j^{\text{cat}}, \boldsymbol{x}_k^{\text{cat}}; \boldsymbol{\nu})$ includes all pairwise interactions between one element of $\boldsymbol{X}_j^{\text{cat}}$ and one element of $\boldsymbol{X}_k^{\text{cat}}$. This allows correlation between $\boldsymbol{X}^{\text{cat}}$ of members of the same matched set. Also assume that for $j = 1, \ldots, M+1$ independently,

$$\boldsymbol{X}_j^{\text{con}} \mid D_1 = 1, D_2 = \ldots = D_{M+1} = 0, \boldsymbol{X}_1^{\text{cat}}, \boldsymbol{X}_2^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}}, \boldsymbol{u} \sim N(\boldsymbol{\eta} + \boldsymbol{\xi} I(j=1) + \boldsymbol{\rho} \boldsymbol{X}_j^{\text{cat}} + \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}} + \boldsymbol{u}, \boldsymbol{\Lambda}) \tag{8}$$

$$\text{and } \boldsymbol{u} \mid D_1 = 1, D_2 = \ldots = D_{M+1} = 0, \boldsymbol{X}_1^{\text{cat}}, \boldsymbol{X}_2^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}) \tag{9}$$

where $\bar{\boldsymbol{X}}^{\text{cat}} = (M+1)^{-1} \sum_{j=1}^{M+1} \boldsymbol{X}_j^{\text{cat}}$. Note that $\boldsymbol{\psi}$ and $\boldsymbol{\Omega}$ allow correlation between one individual's $\boldsymbol{X}^{\text{con}}$ and the $\boldsymbol{X}^{\text{cat}}$ and $\boldsymbol{X}^{\text{con}}$ of other members of the same matched set. In Web

Appendix E, we show this model implies equation (2) holds with $\boldsymbol{\beta}_{\text{cat}} = \boldsymbol{\tau} - \boldsymbol{\rho}^\top(\boldsymbol{F} - \boldsymbol{C})\boldsymbol{\xi}$ and $\boldsymbol{\beta}_{\text{con}} = (\boldsymbol{F} - \boldsymbol{C})\boldsymbol{\xi}$, where $\boldsymbol{C}^{-1} = \boldsymbol{\Lambda} + \boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\Lambda} + M\boldsymbol{\Omega})^{-1}M\boldsymbol{\Omega}$ and $\boldsymbol{F} = -(\boldsymbol{\Lambda} + M\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}\boldsymbol{C}$.

As in Section 4, we propose using FCS MI with conditional models compatible with this joint model. In Web Appendix F, we show that a compatible conditional model for a partially observed element of $\boldsymbol{X}_j^{\text{con}}$ is a linear regression of that element on $\boldsymbol{X}_j^{\text{cat}}$, $\sum_{k\neq j}\boldsymbol{X}_k^{\text{cat}}$, $\sum_{k\neq j}\boldsymbol{X}_k^{\text{con}}$ and all the remaining elements of $\boldsymbol{X}_j^{\text{con}}$. Likewise, a compatible conditional model for one of the partially observed categorical variables making up $\boldsymbol{X}_j^{\text{cat}}$ is a multinomial logistic regression of this categorical variable on $\boldsymbol{X}_j^{\text{con}}$, $\sum_{k\neq j}\boldsymbol{X}_k^{\text{con}}$, $\sum_{k\neq j}\boldsymbol{X}_k^{\text{cat}}$, and those elements of $\boldsymbol{X}_j^{\text{cat}}$ that are not dummy indicators for this categorical variable. These conditional models are not the default options in MI software, because some predictors in the regression are sums of conditioning variables, e.g. $\sum_{k\neq j}\boldsymbol{X}_k^{\text{cat}}$. However, specification of non-default conditional models is straightforward (see Web Appendix H).

As with the FCS method in Section 4, this method is asymptotically equivalent to joint model MI, but in finite samples may be inefficient. Our second proposed model for $\boldsymbol{X}^{\text{set}}$ is a latent normal model with random effects (Carpenter and Kenward, 2013). Like the latent normal model of Section 4, this is not compatible with equation (2), but its use may improve efficiency. The latent normal model is the same as model (5), but with $\boldsymbol{\delta}_{\text{LN}}\boldsymbol{S}$ replaced by $\boldsymbol{u}$ and now conditioning on all of $D_1, \ldots, D_{M+1}$: for $j = 1, \ldots, M+1$ independently,

$$(\boldsymbol{X}_j^{\text{con}\top}, \boldsymbol{W}_j^{\text{cat}\top})^\top \mid D_1 = 1, D_2 = \ldots = D_{M+1} = 0, \boldsymbol{u} \sim N(\boldsymbol{\alpha}_{\text{LN}} + \boldsymbol{\phi}_{\text{LN}}D_j + \boldsymbol{u},\ \boldsymbol{\Sigma}_{\text{LN}}) \quad (10)$$

where $\boldsymbol{u}$ is normally distributed with mean zero and unstructured variance given $D_1 = 1, D_2 = \ldots = D_{M+1} = 0$. Again, joint model MI can be done using REALCOM-MI or jomo.

As in Section 4, there is a normal version of this model. This assumes (10) but with $\boldsymbol{W}_j^{\text{cat}}$ replaced by $\boldsymbol{X}_j^{\text{cat}}$. Joint model MI with this model can be done using the pan package of R.

6. Simulation Study

One thousand datasets were generated for each of 24 scenarios resulting from considering two sample sizes ($N = 100$ or 500 matched sets), two numbers of matching controls ($M = 1$ or $M = 4$), three missingness mechanisms, and two proportions of missing data. Each dataset was generated using the model defined by expressions (3)–(4). Specifically, there were two matching variables, one binary (S^{cat}) and one continuous (S^{con}), and three covariates, one categorical (X^{cat}) and two continuous (X^{conA} and X^{conB}). We assumed $P(S^{\text{cat}} = 1 \mid D = 1) = 0.6$ and $S^{\text{con}} \mid S^{\text{cat}}, D = 1 \sim N(0, 1)$. These could represent, for example, sex and standardised age. Among cases, the sex with greater risk would be more common, while age might be approximately normal if risk increases with age but total population size diminishes due to all-cause mortality. We assumed logit $P(X^{\text{cat}} = 1 \mid S^{\text{cat}}, S^{\text{con}}, D) = -2.5 + 0.5S^{\text{cat}} + 0.5S^{\text{con}} + 0.75D$ (so about 10% of controls and 20% of cases have $X^{\text{cat}} = 1$), and $(X^{\text{conA}}, X^{\text{conB}})$ given $X^{\text{cat}}, S^{\text{cat}}, S^{\text{con}}, D$ is bivariate normal with univariate marginal distributions $N(0.5X^{\text{cat}} + 0.5S^{\text{cat}} + 0.5S^{\text{con}} + 0.5D, 1)$ and covariance 0.5. From $\boldsymbol{\beta}_{\text{cat}} = \boldsymbol{\lambda} - \boldsymbol{\gamma}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}$ and $\boldsymbol{\beta}_{\text{con}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}$, the true log ORs of X^{cat}, X^{conA} and X^{conB} are $\beta_{\text{cat}} = 5/12$, $\beta_{\text{conA}} = 1/3$ and $\beta_{\text{conB}} = 1/3$.

Missingness was imposed on X^{cat} and X^{conA} assuming either missing completely at random (MCAR) or one of two MAR mechanisms. For MCAR data, each individual's X^{cat} and X^{conA} variables were independently missing with probability p_{miss}. Two values, $p_{\text{miss}} = 0.1$ and $p_{\text{miss}} = 0.25$, were considered. Thus, either 19% or 44% of individuals had at least one missing variable. For the first MAR mechanism (MAR-A), each individual's X^{cat} and X^{conA} variables were independently missing with logit probability $c_{\text{miss}} + 0.25(X^{\text{conB}} + S^{\text{cat}} + S^{\text{con}} + D)$. For the second (MAR-B), it was $c_{\text{miss}} + 0.25(X^{\text{conB}} + S^{\text{cat}} + S^{\text{con}} + D + X^{\text{conB}}D)$. In both cases, c_{miss} was chosen to give $p_{\text{miss}} = 0.1$ or $p_{\text{miss}} = 0.25$ missingness in each of X^{cat} and X^{conA}.

Each dataset was analyzed using CLR with X^{cat}, X^{conA} and X^{conB} as covariates. Missing

data were handled in seven ways: complete-case analysis; FCS MI using matching variables or matched set (using ice in Stata); latent normal MI using matching variables or matched set (jomo in R); and normal MI using matching variables (mi impute in Stata) or matched set (pan in R). We used 25 imputed datasets when $p_{\text{miss}} = 0.1$, and 50 when $p_{\text{miss}} = 0.25$. In addition, the complete data were analyzed before imposing missingness on the covariates.

Tables 1 and 2 show results for the MCAR mechanism with 1:1 matching ($M = 1$) and 1:4 matching ($M = 4$) when $N = 500$. The results from these scenarios also give a good indication of the general patterns observed for MAR-A, MAR-B and $N = 100$ (see Web Tables 1–10). We shall focus on β_{cat} and β_{conA}, since for β_{conB} (the fully-observed covariate), differences between the three MI methods using matched variables were small, as were differences between those using matched set. To ease comparison of the six MI methods, Tables 3 and 4 show, for $p_{\text{miss}} = 0.25$, the bias, ratio of empirical SEs, ratio of mean estimated SEs, and relative efficiency (i.e. ratio of mean squared errors, MSE) of each method, averaged over the three missingness mechanisms, separately for β_{cat} and β_{conA}, for $N = 100$ and 500, and for $M = 1$ and $M = 4$. Unsurprisingly, differences between methods were smaller when $p_{\text{miss}} = 0.1$; we focus on $p_{\text{miss}} = 0.25$ below.

The two FCS methods are approximately unbiased when $N = 500$ and usually when $N = 100$. Exceptions are when $N = 100$ and $M = 1$, where the complete-data method is also biased (with biases similar to those of FCS MI), and when $N = 100$ and $M = 4$, where there is bias for β_{cat} when using matched set. Normal MI has some negative bias for β_{cat}, especially when using matching variables (except when $N = 100$ and $M = 1$, where its negative bias cancels out the positive bias of the complete-data estimator). Latent normal MI has some positive bias for β_{cat} when $M = 1$; latent normal MI using matched set also has negative bias for β_{conA}. The complete-case estimators are generally approximately unbiased, but note that the estimator of β_{conB} is severely biased under MAR-B (Web Tables 2, 4, 7 and 10).

Empirical standard errors (SEs) from MI are almost always smaller when using matching variables than when using matched set, and negatively biased estimators tend to have smaller SEs. For β_{cat}, the SEs from FCS MI and latent normal MI are usually similar (when using matched set with $N = 100$, FCS MI has the smaller SE); the smallest SEs come from normal MI. For β_{conA}, latent normal MI has the smallest SEs; the SEs from normal MI are similar to those from FCS MI when using matching variables and larger when using matched set. These differences are less marked when $M = 4$.

Efficiency (mean square error, MSE) is a function of bias and SE. For β_{cat}, normal MI is most efficient, despite its bias; FCS MI and latent normal MI are usually about equally efficient, with neither uniformly better than the other. For β_{conA}, latent normal MI is more efficient than FCS and normal MI when using matching variables; FCS and normal MI are equally efficient. When using matched set, FCS MI is more efficient for β_{conA} than normal MI; latent MI is more efficient than FCS MI when $M = 1$, but is the least efficient of all the methods when $N = 500$ and $M = 4$, where its bias dominates its smaller SE. All MI methods are more efficient than the complete-case analysis.

The MI methods show a tendency to slightly overestimate SEs. Mostly, this is fairly mild, but is more severe for normal MI with β_{cat} when $M = 1$, and for latent normal MI with β_{conA} when $M = 1$ or 4. Thus, although normal and latent normal MI are most efficient for β_{cat} and β_{conA}, respectively, this advantage is not apparent in the width of the estimated confidence intervals. Indeed, the average estimated SEs of the three MI methods using matching variables were generally rather similar; the same was true of the methods using matched set. Coverage of 95% confidence intervals was between 93% and 97% for all methods.

We also performed two simulation studies using modified data-generating mechanisms that make our imputation models misspecified. In the first, there was an interaction between S^{cat} and S^{con}; in the second, X^{conA} and X^{conB} were log-normally distributed. See Web Appendix G

and Web Tables 11–16 for details and results. Briefly, none of the MI methods showed considerable bias for either of these data-generating mechanisms, and all MI methods were much more efficient than the complete-case analysis.

In summary, all the MI methods appear to work well. Using matching variables is more efficient than using matched set. If using matching variables, normal and latent normal MI appear to be preferable to FCS MI, which is less efficient; normal MI is more efficient for β_{cat}, but latent normal MI more efficient for β_{conA}. Of these, one might prefer latent normal MI, because of the bias in β_{cat} for normal MI. If using matched set, FCS MI might be preferred when $M = 4$, on bias and efficiency grounds. However, when $M = 1$ and using matched set, no method appears better than any other.

7. Analysis of EPIC-Norfolk Data

Table 5 shows the estimated adjusted log OR for fibre intake from the complete-case analysis. This analysis excludes all matched sets in which the case had missing data, as well as any controls with missing data. It uses 240 (75%) matched sets, consisting of 240 cases and 772 controls. Also shown are the results of the three MI methods using matching variables, including sex, age and date of diary completion as $\boldsymbol{S}$ in the imputation model. The complete-case and MI analyses produce similar log OR estimates (differing by less than 20% of a SE), but the latter are more efficient, because they use all 318 cases and 1272 controls i.e. 33% more matched sets, and this is reflected by a 17% reduction in estimated SE.

MI using matching variables imputes missing values assuming that age, sex and time of diary completion have linear and additive effects on the logit probability of disease. Furthermore, the way that recruitment took place in the EPIC-Norfolk cohort means that date of diary completion is predictive of which GP surgery the individual was registered with, and hence matching by the former tends to induce some degree of matching by the latter. Treating

date of diary completion as a continuous variable will not fully account for this. For these two reasons, one might prefer MI using matched set, or might wish to check that the results from the two methods do not differ substantially.

Table 5 shows that the results from MI using matching variables and MI using matched set are very similar, providing some reassurance about the validity of both sets of results. In this study, both approaches can be used, but had matching been on GP practice itself, MI using matched set might have been the only feasible option.

8. Discussion

We have described two broad MI approaches to the analysis of matched case-control studies with missing values in covariates, and three methods within each approach. One approach involves parametric modelling of the association between the matching variables and the partially observed covariates; the other instead treats matched set as a random effect. Our simulation results suggest that the first approach is preferable when it can be done, as it is more efficient. However, in studies where matching is on, e.g. family, GP practice or postcode area of residence, or if data on the matching variables are not available to the analyst, the first approach is not feasible and the second approach can be used instead. The second approach might also be preferred if one were reluctant to specify a form for association between matching variables and covariates in the imputation model, because, for example, there were several matching variables, including continuous ones and potential interactions.

Of the three MI methods within each approach, FCS MI based on a restricted general location model and joint model MI using a multivariate normal distribution can be implemented in many statistical packages, whereas joint model MI using a latent normal distribution is currently limited to R and the specialist software REALCOM-MI. All three methods are easy to use, appear to work well, and are more efficient than the complete-case analysis. They can all handle continuous and nominal categorical covariates, multiple partially-observed

covariates, non-monotone missingness patterns and multiple controls per case. Computer commands to implement the methods are given in Web Appendix H.

FCS MI has the theoretical appeal of being asymptotically equivalent to joint model MI using an imputation model (the restricted general location model) that is compatible with the CLR analysis model. It nearly always gave the least biased estimates in simulations. However, when using matching variables, normal MI and latent normal MI were more efficient. When using matched set, FCS MI was marginally better than normal and latent normal MI for a 1:4 matched study ($M = 4$); no method was obviously best or worst for 1:1 matching ($M = 1$). A drawback of FCS MI using matched set when $M > 1$ is that the estimates may depend on the arbitrary order chosen for the M controls in each matched set. Any order produces valid imputations, but one could avoid this dependence by randomly permuting indices of controls within matched sets before generating each imputed dataset. However, that is only likely to be worthwhile if the sample size is small and there is a lot of missing data. Normal MI was, in general, the most biased of the three methods, but even its biases were fairly modest. A slight drawback of normal MI is the need manually to post-process imputed values of categorical variables, e.g. using adaptive rounding. None of the MI methods was uniformly superior to the others in simulations, and we regard use of any of them as entirely acceptable.

All methods can handle the situation where the number of matched controls, M, varies between cases, although this is slightly more complicated for FCS MI using matched set. For this method, extra controls with completely missing data would have to be added to those matched sets with fewer than the maximum number of controls, before performing MI, and then deleted again before analysing the imputed datasets.

Another method, which merits further research, is joint model MI using the restricted general location model. This requires specialist Bayesian software and more advanced programming skills, and the focus of this article is on methods that are easy to implement in standard pack-

ages. Nevertheless, it would be worth investigating whether this method is significantly more efficient than our FCS MI method based on the same model. The mix package in R (Schafer, 1997) may also be of interest. This uses a model similar to (3)–(4), but additionally assumes the continuous part of $\boldsymbol{S}$ is normally distributed given D, $\boldsymbol{X}^{\text{cat}}$ and the rest of $\boldsymbol{S}$. The MI methods considered in this article assume, like the CLR analysis model, nothing about the distribution of the matching variables. Mix cannot be used for MI using matched set.

Finally, we note that, as always with missing data methods, it is important to consider the plausibility of the assumption about the missing data mechanism. Often, the MAR assumption can be made more plausible by including in the imputation model additional variables that are associated with the partially observed covariates.

Supplementary Materials

Web Appendices referenced in Sections 3–8, along with computer code, are available with this paper at the Biometrics website on Wiley Online Library.

Acknowledgements

SRS is funded by MRC grants U105260558 and MC_US_A030_0015. We would like to acknowledge the contribution of the staff and participants of the EPIC-Norfolk Study. EPIC-Norfolk is supported by the MRC programme grant (G0401527).

References

Ahn, J., Mukherjee, B., Gruber, S., and Sinha, S. (2011). Missing exposure data in stereotype regression model: Application to matched case-control study with disease subclassification. *Biometrics* **67,** 546–558.

Asano, T. and McLeod, R. (2004). Non steroidal anti-inflammatory drugs (NSAID) and aspirin for preventing colorectal adenomas and carcinomas. *Cochrane Database of Systematic Reviews* **Issue 1,** Art. No.: CD004079. DOI: 10.1002/14651858.CD004079.pub2.

Aune, D., Chan, D., Lau, R., Vieira, R., Greenwood, D., Kampman, E., and Norat, T.

(2011). Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *British Medical Journal* **343,** d6617.

Bernaards, C., Belin, T., and Schafer, J. (2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Statistics in Medicine* **26,** 1368–82.

Breslow, N. and Day, N. (1980). *Statistical Methods in Cancer Research, Volume I – The analysis of case-control studies*. IARC Scientific Publications.

Carpenter, J. and Kenward, M. (2013). *Multiple Imputation and its Applications*. Wiley, New Jersey.

Dahm, C., Keogh, R., Spencer, E., Greenwood, D., Key, T., Fentiman, I., Shipley, M., Brunner, E., Cade, J., Burley, V., Mishra, G., Stephen, A., Kuh, D., White, I., Luben, R., Lentjes, M., Khaw, K., and Rodwell, S. (2010). Dietary fiber and colorectal cancer risk: a nested case-control study using food diaries. *Journal of the National Cancer Institute* **102,** 614–626.

Gebregziabher, M. and DeSantis, S. (2010). Latent class based multiple imputation approach for missing categorical data. *Journal of Statistical Planning and Inference* **140,** 3252–62.

Hughes, R., White, I., Seaman, S., Carpenter, J., Tilling, K., and Sterne, J. (2014). Joint modelling rationale for chained equations. *BMC Medical Research Methodology* **14,** 28.

Lee, K. and Carlin, J. (2010). Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology* **171,** 624–632.

Lipsitz, S., Parzen, M., and Ewell, M. (1998). Inference using conditional logistic regression with missing covariates. *Biometrics* **54,** 295–303.

Liu, J., Gelman, A., Hill, J., Su, Y., and Kropko, J. (2014). On the stationary distribution of iterative imputations. *Biometrika* **101,** 155–173.

Liu, T., Yuan, X., Li, Z., and Li, Y. (2013). Empirical and weighted conditional likelihoods for matched case-control studies with missing covariates. *Journal of Multivariate Analysis* **119,** 185–199.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* **10,** 325–337.

Moons, K., Donders, R., Stijnen, T., and Harrell, F. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology* **59,** 1092–1101.

National Cancer Institute (2014). Colorectal cancer prevention (PDQ). http://www.cancer.gov/cancertopics/pdq/prevention/colorectal/HealthProfessional/page3. Accessed: 17/06/2014.

Paik, M. (2004). Nonignorable missingness in matched case-control data analyses. *Biometrics* **60,** 306–314.

Paik, M. and Sacco, R. (2000). Matched case-control data analyses with missing covariates. *Applied Statistics* **49,** 145–156.

Rathouz, P. (2003). Likelihood methods for missing covariate data in highly stratified studies. *Journal of the Royal Statistical Society, Series B* **65,** 711–723.

Rathouz, P., Satten, G., and Carroll, R. (2002). Semiparametric inference in matched casecontrol studies with missing covariate data. *Biometrika* **89,** 905–916.

Satten, G. and Carroll, R. (2000). Conditional and unconditional categorical regression models with missing covariates. *Biometrics* **56,** 384–388.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.

Sinha, S., Mukherjee, B., Ghosh, M., Mallick, B., and Carroll, R. (2005). Semiparametric Bayesian analysis of matched case-control studies with missing exposure. *Journal of the*

American Statistical Association **100,** 591–601.

Sinha, S. and Wang, S. (2009). A new semiparametric procedure for matched case-control studies with missing covariates. *Journal of Nonparametric Statistics* **21,** 889–905.

van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Chapman & Hall/CRC.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

Web-based Supplementary Materials for 'Handling Missing Data in Matched Case-Control Studies using Multiple Imputation' by SR Seaman and RH Keogh

Web Appendix A: Proof that equation (1) implies (2) and vice versa

First, suppose that equation (1) holds for any function $q(\boldsymbol{S})$. Then

$$\begin{aligned}
&G(\boldsymbol{x}_1^{\text{cat}}, \boldsymbol{x}_1^{\text{con}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}}, \boldsymbol{x}_{M+1}^{\text{con}}) \\
&= \frac{P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}_1^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}_1^{\text{con}}, \boldsymbol{S}) \prod_{k=2}^{M+1} P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}_k^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}_k^{\text{con}}, \boldsymbol{S})}{\sum_{j=1}^{M+1} P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}_j^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}_j^{\text{con}}, \boldsymbol{S}) \prod_{k\neq j} P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}_k^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}_k^{\text{con}}, \boldsymbol{S})} \\
&= \frac{\exp(\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}_1^{\text{con}})}{\sum_{j=1}^{M+1} \exp(\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}_j^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}_j^{\text{con}})}
\end{aligned}$$

Second, suppose that equation (2) holds. Then

$$G(\boldsymbol{x}^{\text{cat}}, \boldsymbol{x}^{\text{con}}, \boldsymbol{0}, \boldsymbol{0}, \ldots, \boldsymbol{0}, \boldsymbol{0}) = \frac{\exp(\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}^{\text{con}})}{\exp(\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}^{\text{con}}) + M}.$$

Therefore

$$\begin{aligned}
&\frac{\exp(\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}^{\text{con}})}{\exp(\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}^{\text{con}}) + M}. \\
&= \frac{P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}^{\text{con}}, \boldsymbol{S}) P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S})^M}{\left\{ \begin{array}{l} P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}^{\text{con}}, \boldsymbol{S}) P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S})^M \\ +M P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S}) P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}^{\text{con}}, \boldsymbol{S}) \\ \quad \times P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S})^{M-1} \end{array} \right\}}.
\end{aligned}$$

This implies that

$$\begin{aligned}
&M \exp(-\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}^{\text{cat}} - \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}^{\text{con}}) + 1 \\
&= 1 + M \frac{P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S}) P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}^{\text{con}}, \boldsymbol{S})}{P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}^{\text{con}}, \boldsymbol{S}) P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S})}
\end{aligned}$$

which, in turn, implies that

$$\begin{aligned}
&\exp(\boldsymbol{\beta}_{\text{cat}}^\top \boldsymbol{x}^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^\top \boldsymbol{x}^{\text{con}}) \\
&= \frac{P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}^{\text{con}}, \boldsymbol{S})}{P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{x}^{\text{cat}}, \boldsymbol{X}^{\text{con}}=\boldsymbol{x}^{\text{con}}, \boldsymbol{S})} \times \frac{P(D=0 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S})}{P(D=1 \mid \boldsymbol{X}^{\text{cat}}=\boldsymbol{0}, \boldsymbol{X}^{\text{con}}=\boldsymbol{0}, \boldsymbol{S})}.
\end{aligned}$$

Since this is true for any $\boldsymbol{x}^{\text{cat}}$, $\boldsymbol{x}^{\text{con}}$ and any value of $\boldsymbol{S}$, equation (1) holds with $q(\boldsymbol{S}) =$ logit $P(D = 1 \mid \boldsymbol{X}^{\text{cat}} = \boldsymbol{0}, \boldsymbol{X}^{\text{con}} = \boldsymbol{0}, \boldsymbol{S})$.

Web Appendix B: Joint Model MI and Full-Conditional Specification (FCS) MI

In this appendix, we provide a more technical description of joint model MI and FCS MI than that given in the main text of the paper. We begin with joint model MI, and then discuss FCS MI and its relation to joint model MI.

Consider a generic dataset with n units and K random variables $Y_1, \ldots, Y_K$ potentially measured on each unit. Let $\boldsymbol{Y}_{-k} = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_K)$. Use subscript i to index the unit. So, Y_{ik} $(i = 1, \ldots, n)$ and $\boldsymbol{Y}_{i,-k}$ denote, respectively, Y_k and $\boldsymbol{Y}_{-k}$ for the ith unit. Let $\boldsymbol{Y}_{\bullet k} = (Y_{1k}, \ldots, Y_{nk})$ and $\boldsymbol{Y}_{\bullet -k} = (\boldsymbol{Y}_{\bullet 1}, \ldots, \boldsymbol{Y}_{\bullet k-1}, \boldsymbol{Y}_{\bullet k+1}, \ldots, \boldsymbol{Y}_{\bullet K})$. Let $\boldsymbol{Y}_{\bullet k}^{\text{obs}}$ and $\boldsymbol{Y}_{\bullet k}^{\text{mis}}$ denote, respectively, the observed and missing parts of $\boldsymbol{Y}_{\bullet k}$, and let $\boldsymbol{Y}_{\bullet -k}^{\text{obs}}$ and $\boldsymbol{Y}_{\bullet -k}^{\text{mis}}$ denote the observed and missing parts of $\boldsymbol{Y}_{\bullet -k}$.

In joint model MI, a model $f(Y_1, \ldots, Y_K \mid \boldsymbol{\theta})$ is specified for $Y_1, \ldots, Y_K$, with prior distribution $\pi(\boldsymbol{\theta})$ on the parameters $\boldsymbol{\theta}$ of this model. Random vectors $(Y_{i1}, \ldots, Y_{iK})$ and $(Y_{j1}, \ldots, Y_{jK})$ $(i \neq j)$ are assumed to be conditionally independent given $\boldsymbol{\theta}$. Let $f_k(Y_k \mid \boldsymbol{Y}_{-k};\ \boldsymbol{\theta})$ denote the full-conditional distribution of Y_k, i.e. the distribution of Y_k given $\boldsymbol{Y}_{-k}$ and $\boldsymbol{\theta}$ implied by $f(Y_1, \ldots, Y_K \mid \boldsymbol{\theta})$. Imputed values for $\boldsymbol{Y}_{\bullet 1}^{\text{mis}}, \ldots, \boldsymbol{Y}_{\bullet K}^{\text{mis}}$ are drawn from their posterior predictive distribution:

$$p(\boldsymbol{Y}_{\bullet 1}^{\text{mis}}, \ldots, \boldsymbol{Y}_{\bullet K}^{\text{mis}} \mid \boldsymbol{Y}_{\bullet 1}^{\text{obs}}, \ldots, \boldsymbol{Y}_{\bullet K}^{\text{obs}}) \ \propto \int f(\boldsymbol{Y}_{\bullet 1}^{\text{mis}}, \ldots, \boldsymbol{Y}_{\bullet K}^{\text{mis}} \mid \boldsymbol{Y}_{\bullet 1}^{\text{obs}}, \ldots, \boldsymbol{Y}_{\bullet K}^{\text{obs}}, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta}) \times f(\boldsymbol{Y}_{\bullet 1}^{\text{obs}}, \ldots, \boldsymbol{Y}_{\bullet K}^{\text{obs}} \mid \boldsymbol{\theta})\, d\boldsymbol{\theta}$$

(Little and Rubin, 2002). This can be achieved by using a Gibbs sampler algorithm. One form of the Gibbs sampler is as follows (Liu et al., 2014). Initially, the missing values of Y_k $(k = 1, \ldots, K)$ are replaced by the mean, or a random sample, of the observed values of Y_k.

A single iteration of the Gibbs sampler involves K steps, in the kth of which the missing values of Y_k are updated. The kth step ($k = 1, \ldots, K$) of the algorithm is to sample $\boldsymbol{\theta}$ from its posterior distribution given $\boldsymbol{Y}_{\bullet k}^{\text{obs}}$ and $\boldsymbol{Y}_{\bullet -k}$ (using the current value of $\boldsymbol{Y}_{\bullet -k}^{\text{mis}}$) and then, for each individual i ($i = 1, \ldots, n$) with missing Y_{ik}, to sample Y_{ik} from $f_k(Y_{ik} \mid \boldsymbol{Y}_{i,-k};\ \boldsymbol{\theta})$ using the sampled value of $\boldsymbol{\theta}$. Step k is omitted if Y_k is fully observed. The Gibbs sampler is iterated until convergence, at which point $\boldsymbol{Y}_{\bullet 1}^{\text{mis}}, \ldots, \boldsymbol{Y}_{\bullet K}^{\text{mis}}$ are drawn from their posterior predictive distribution. This algorithm is applied repeatedly to the original dataset, in order to create multiple imputed datasets. The analysis model is then fitted to each imputed dataset separately and the resulting parameter and variance estimates combined using Rubin's Rules (Little and Rubin, 2002). The imputation model is said to be *compatible with the analysis model* if there exists a model for the joint distribution of all the variables that implies the analysis and imputation models as submodels. When the imputation model is correctly specified and compatible with the analysis model, and data are MAR, joint model MI gives consistent parameter and variance estimates for the analysis model (Little and Rubin, 2002).

An alternative to joint model MI is FCS MI (also known as MI by chained equations) (van Buuren, 2012). FCS MI avoids the need to specify a model for the joint distribution of $Y_1, \ldots, Y_K$ and to sample from the corresponding posterior of $\boldsymbol{\theta}$ given $\boldsymbol{Y}_{\bullet k}^{\text{obs}}$ and $\boldsymbol{Y}_{\bullet -k}$ and from the full-conditionals. Instead, for each $k = 1, \ldots, K$, a model $g_k(Y_k \mid \boldsymbol{Y}_{-k};\ \boldsymbol{\theta}_k)$ is specified for the conditional distribution of Y_k given $\boldsymbol{Y}_{-k}$, and a non-informative prior $\pi_k(\boldsymbol{\theta}_k)$ is specified for the parameters $\boldsymbol{\theta}_k$ of this model. FCS proceeds as per the Gibbs sampler except that at step k, $\boldsymbol{\theta}_k$ is sampled from the posterior distribution proportional to the likelihood formed by the product of $g_k(Y_{ik} \mid \boldsymbol{Y}_{i,-k};\ \boldsymbol{\theta}_k)$ over units with observed Y_{ik} multiplied by the prior $\pi_k(\boldsymbol{\theta}_k)$, and missing values of Y_{ik} are sampled from $g_k(Y_{ik} \mid \boldsymbol{Y}_{i,-k};\ \boldsymbol{\theta}_k)$. Step k and the specification of $g_k(Y_k \mid \boldsymbol{Y}_{-k};\ \boldsymbol{\theta}_k)$ is omitted if Y_k is fully observed.

An important theoretical result about the asymptotic relation between joint model MI and

FCS MI was provided by Liu et al. (2014). They defined the set of conditional models $\{g_k(Y_k \mid \boldsymbol{Y}_{-k};\ \boldsymbol{\theta}_k) :\ k = 1, \ldots, K\}$ to be 'compatible' with a joint model $f(Y_1, \ldots, Y_K;\ \boldsymbol{\theta})$ if the following condition holds for each $k = 1, \ldots, K$. For each value of $\boldsymbol{\theta}_k$ in its parameter space, there exists at least one value of $\boldsymbol{\theta}$ in its parameter space for which $g_k(Y_k \mid \boldsymbol{Y}_{-k};\ \boldsymbol{\theta}_k) = f_k(Y_k \mid \boldsymbol{Y}_{-k};\ \boldsymbol{\theta})$. (Note that there is no need for there to exist a value of $\boldsymbol{\theta}$ for which this is true for all $k = 1, \ldots, K$ simultaneously.) Liu et al. (2014) showed that when the set of conditional models is compatible with a joint model, the distribution of the imputed data converges, as the sample size tends to infinity, to the posterior predictive distribution of the missing data under that joint model. Hence, the use of FCS MI is asymptotically valid in this case.

For more information about joint model MI and FCS MI, refer to, for example, Carpenter and Kenward (2013).

Web Appendix C: Proof that equations (3) and (4) imply that (2) holds with $\boldsymbol{\beta}_{\text{con}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}$ and $\boldsymbol{\beta}_{\text{cat}} = \boldsymbol{\lambda} - \boldsymbol{\gamma}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}$

$$\log p(\boldsymbol{X}^{\text{con}}, \boldsymbol{X}^{\text{cat}} \mid \boldsymbol{S}, D) = \log p(\boldsymbol{X}^{\text{con}} \mid \boldsymbol{X}^{\text{cat}}, \boldsymbol{S}, D) + \log p(\boldsymbol{X}^{\text{cat}} \mid \boldsymbol{S}, D) \tag{11}$$

From equation (4),

$$\begin{aligned}
&\log p(\boldsymbol{X}^{\text{con}} \mid \boldsymbol{X}^{\text{cat}}, \boldsymbol{S}, D = 1) - \log p(\boldsymbol{X}^{\text{con}} \mid \boldsymbol{X}^{\text{cat}}, \boldsymbol{S}, D = 0) \\
&= \boldsymbol{\phi}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}^{\text{con}} - \boldsymbol{\alpha} - \boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} - \boldsymbol{\delta}\boldsymbol{S}) - \boldsymbol{\phi}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}/2 \\
&= \boldsymbol{\phi}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}^{\text{con}} - \boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}}) - \boldsymbol{\phi}^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\alpha} + \boldsymbol{\delta}\boldsymbol{S}) - \boldsymbol{\phi}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}/2
\end{aligned} \tag{12}$$

Note that only the first term in expression (12) depends on $\boldsymbol{X}^{\text{cat}}$ or $\boldsymbol{X}^{\text{con}}$.

It follows from equation (3) that

$$\begin{aligned}
&\log P(\boldsymbol{X}^{\text{cat}} = \boldsymbol{x}^{\text{cat}} \mid \boldsymbol{S}, D = 1) - \log P(\boldsymbol{X}^{\text{cat}} = \boldsymbol{x}^{\text{cat}} \mid \boldsymbol{S}, D = 0) \\
&= \boldsymbol{\lambda}^\top\boldsymbol{x}^{\text{cat}} + \text{ terms that do not depend on } \boldsymbol{x}^{\text{cat}}.
\end{aligned} \tag{13}$$

From equations (11), (12) and (13), we have,

$$\begin{aligned}
&\log p(\boldsymbol{X}^{\text{con}}, \boldsymbol{X}^{\text{cat}} \mid \boldsymbol{S}, D=1) - \log p(\boldsymbol{X}^{\text{con}}, \boldsymbol{X}^{\text{cat}} \mid \boldsymbol{S}, D=0) \\
&= \boldsymbol{\phi}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}^{\text{con}} - \boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}}) + \boldsymbol{\lambda}^{\top}\boldsymbol{X}^{\text{cat}} \\
&\quad + \text{terms that do not depend on } \boldsymbol{X}^{\text{cat}} \text{ or } \boldsymbol{X}^{\text{con}}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\log p(\boldsymbol{X}_1^{\text{con}}, \boldsymbol{X}_1^{\text{cat}} \mid \boldsymbol{S}, D_1=1) - \log p(\boldsymbol{X}_1^{\text{con}}, \boldsymbol{X}_1^{\text{cat}} \mid \boldsymbol{S}, D_1=0) \\
&- \log\left\{\sum_{j=1}^{M+1} p(\boldsymbol{X}_j^{\text{con}}, \boldsymbol{X}_j^{\text{cat}} \mid \boldsymbol{S}, D_j=1) - \log p(\boldsymbol{X}_j^{\text{con}}, \boldsymbol{X}_j^{\text{cat}} \mid \boldsymbol{S}, D_j=0)\right\} \\
&= \boldsymbol{\phi}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_1^{\text{con}} - \boldsymbol{\gamma}\boldsymbol{X}_1^{\text{cat}}) + \boldsymbol{\lambda}^{\top}\boldsymbol{X}_1^{\text{cat}} \\
&\quad - \left\{\sum_{j=1}^{M+1} \boldsymbol{\phi}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_j^{\text{con}} - \boldsymbol{\gamma}\boldsymbol{X}_j^{\text{cat}}) + \boldsymbol{\lambda}^{\top}\boldsymbol{X}_j^{\text{cat}}\right\}
\end{aligned}$$

This does not depend on $\boldsymbol{S}$, and hence

$$G(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{M+1}) = \frac{\exp(\boldsymbol{\beta}_{\text{cat}}^{\top}\boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^{\top}\boldsymbol{x}_1^{\text{con}})}{\sum_{j=1}^{M+1}\exp(\boldsymbol{\beta}_{\text{cat}}^{\top}\boldsymbol{x}_j^{\text{cat}} + \boldsymbol{\beta}_{\text{con}}^{\top}\boldsymbol{x}_j^{\text{con}})}$$

with

$$\begin{aligned}
\boldsymbol{\beta}_{\text{con}} &= \boldsymbol{\Sigma}^{-1}\boldsymbol{\phi} \\
\boldsymbol{\beta}_{\text{cat}} &= \boldsymbol{\lambda} - \boldsymbol{\gamma}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{\phi}
\end{aligned}$$

Web Appendix D: Proof that linear regression and multinomial logistic regression conditional models are compatible with the general location model of equations (3) and (4)

It is obvious from expression (4) that the conditional distribution of any element of $\boldsymbol{X}^{\text{con}}$ given $\boldsymbol{X}^{\text{cat}}$, $\boldsymbol{S}$, D and the remaining elements of $\boldsymbol{X}^{\text{con}}$ is a normal distribution with mean equal to a linear function of those variables and with constant variance. Since the parameters of the joint normal distribution in expression (4) are unconstrained, so are the parameters describing the mean and variance of this conditional distribution. Therefore, a linear regres-

sion of a partially observed element of $\boldsymbol{X}^{\text{con}}$ on $\boldsymbol{X}^{\text{cat}}$, $\boldsymbol{S}$, D and the remaining elements of $\boldsymbol{X}^{\text{con}}$ (with main effects only and no interactions) constitutes a conditional model that is compatible (in the sense of Definition 1 of Liu et al., 2014) with the general location model of equations (3) and (4).

Now consider partially observed categorical covariates. By Bayes' Theorem,

$$
\begin{aligned}
&\log p(\boldsymbol{X}^{\text{cat}} \mid \boldsymbol{X}^{\text{con}}, \boldsymbol{S}, D) \\
&= \log p(\boldsymbol{X}^{\text{cat}} \mid \boldsymbol{S}, D) + \log p(\boldsymbol{X}^{\text{con}} \mid \boldsymbol{X}^{\text{cat}}, \boldsymbol{S}, D) - \log p(\boldsymbol{X}^{\text{con}} \mid \boldsymbol{S}, D) \\
&= a(\boldsymbol{X}^{\text{cat}}, \boldsymbol{S};\, \boldsymbol{\zeta}) + D\boldsymbol{\lambda}^\top \boldsymbol{X}^{\text{cat}} - \frac{1}{2}(\boldsymbol{X}^{\text{con}} - \boldsymbol{\alpha} - \boldsymbol{\phi}D - \boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} - \boldsymbol{\delta}\boldsymbol{S})^\top \boldsymbol{\Sigma}^{-1} \\
&\quad \times (\boldsymbol{X}^{\text{con}} - \boldsymbol{\alpha} - \boldsymbol{\phi}D - \boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} - \boldsymbol{\delta}\boldsymbol{S}) + \text{ terms not involving } \boldsymbol{X}^{\text{cat}} \\
&= a(\boldsymbol{X}^{\text{cat}}, \boldsymbol{S};\, \boldsymbol{\zeta}) + D\boldsymbol{\lambda}^\top \boldsymbol{X}^{\text{cat}} - \frac{1}{2}\boldsymbol{X}^{\text{cat}\top}\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} \\
&\quad + \boldsymbol{X}^{\text{cat}\top}\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}^{\text{con}} - \boldsymbol{\alpha} - \boldsymbol{\phi}D - \boldsymbol{\delta}\boldsymbol{S}) + \text{ terms not involving } \boldsymbol{X}^{\text{cat}} \\
&= a(\boldsymbol{X}^{\text{cat}}, \boldsymbol{S};\, \boldsymbol{\zeta}) - \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} - \boldsymbol{S}^\top \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} + D\left(\boldsymbol{\lambda}^\top - \boldsymbol{\phi}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\right)\boldsymbol{X}^{\text{cat}} \\
&\quad - \frac{1}{2}\boldsymbol{X}^{\text{cat}\top}\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}\boldsymbol{X}^{\text{cat}} + \boldsymbol{X}^{\text{cat}\top}\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{X}^{\text{con}} \\
&\quad + \text{terms not involving } \boldsymbol{X}^{\text{cat}}
\end{aligned} \tag{14}
$$

Now write $\boldsymbol{X}^{\text{cat}} = (\boldsymbol{X}_A^{\text{cat}\top}, \boldsymbol{X}_B^{\text{cat}\top})^\top$, where $\boldsymbol{X}_A^{\text{cat}}$ represents one categorical variable (coded as a vector of dummy indicators), and $\boldsymbol{X}_B^{\text{cat}}$ represents the remaining variables in $\boldsymbol{X}^{\text{cat}}$. Also write

$$
a(\boldsymbol{X}^{\text{cat}}, \boldsymbol{S};\, \boldsymbol{\zeta}) = \boldsymbol{\zeta}_1^\top \boldsymbol{X}_A^{\text{cat}} + \boldsymbol{X}_B^{\text{cat}\top}\boldsymbol{\zeta}_2 \boldsymbol{X}_A^{\text{cat}} + \boldsymbol{S}^\top \boldsymbol{\zeta}_3 \boldsymbol{X}_A^{\text{cat}} + \text{ terms not involving } \boldsymbol{X}_A^{\text{cat}}
$$

To get the distribution of $\boldsymbol{X}_A^{\text{cat}}$ given $\boldsymbol{X}^{\text{con}}$, $\boldsymbol{S}$, D and $\boldsymbol{X}_B^{\text{cat}}$ we take equation (14) and ignore

all terms not involving $\boldsymbol{X}_A^{\text{cat}}$. The result is

$$
\begin{aligned}
&\log p(\boldsymbol{X}_A^{\text{cat}} \mid \boldsymbol{X}_B^{\text{cat}}, \boldsymbol{X}^{\text{con}}, \boldsymbol{S}, D) \\
&= (\boldsymbol{\zeta}_1^\top - \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A) \boldsymbol{X}_A^{\text{cat}} + \boldsymbol{X}_B^{\text{cat}\top} \boldsymbol{\zeta}_2 \boldsymbol{X}_A^{\text{cat}} + \boldsymbol{S}^\top (\boldsymbol{\zeta}_3 - \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A) \boldsymbol{X}_A^{\text{cat}} \\
&\quad + D(\boldsymbol{\lambda}_A^\top - \boldsymbol{\phi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A) \boldsymbol{X}_A^{\text{cat}} \\
&\quad - \frac{1}{2} \boldsymbol{X}_A^{\text{cat}\top} \boldsymbol{\gamma}_A^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A \boldsymbol{X}_A^{\text{cat}} - \boldsymbol{X}_B^{\text{cat}\top} \boldsymbol{\gamma}_B^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A \boldsymbol{X}_A^{\text{cat}} + \boldsymbol{X}_A^{\text{cat}\top} \boldsymbol{\gamma}_A^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X}^{\text{con}} \\
&\quad + \text{ terms not involving } \boldsymbol{X}_A^{\text{cat}} \\
&= \left(\boldsymbol{\zeta}_1^\top - \boldsymbol{\alpha}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A - \frac{1}{2} \text{diag}(\boldsymbol{\gamma}_A^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A)^\top \right) \boldsymbol{X}_A^{\text{cat}} + \boldsymbol{X}_B^{\text{cat}\top} (\boldsymbol{\zeta}_2 - \boldsymbol{\gamma}_B^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A) \boldsymbol{X}_A^{\text{cat}} \\
&\quad + \boldsymbol{S}^\top (\boldsymbol{\zeta}_3 - \boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A) \boldsymbol{X}_A^{\text{cat}} + D(\boldsymbol{\lambda}_A^\top - \boldsymbol{\phi}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A) \boldsymbol{X}_A^{\text{cat}} \\
&\quad + \boldsymbol{X}^{\text{con}\top} \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A \boldsymbol{X}_A^{\text{cat}} \\
&\quad + \text{ terms not involving } \boldsymbol{X}_A^{\text{cat}}
\end{aligned} \tag{15}
$$

where $\boldsymbol{\gamma}_A$, $\boldsymbol{\gamma}_B$, $\boldsymbol{\lambda}_A$ and $\boldsymbol{\lambda}_B$ are given by the partitions $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_A \ \ \boldsymbol{\gamma}_B]$ and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_A^\top, \boldsymbol{\lambda}_B^\top)^\top$. Note that $\boldsymbol{X}_A^{\text{cat}\top} \boldsymbol{\gamma}_A^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A \boldsymbol{X}_A^{\text{cat}} = \text{diag}(\boldsymbol{\gamma}_A^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_A)^\top \boldsymbol{X}_A^{\text{cat}}$ because $\boldsymbol{X}_A^{\text{cat}}$ is a vector whose elements all equal zero except for at most one element.

It follows from equation (15) that the distribution of $\boldsymbol{X}_A^{\text{cat}}$ given $\boldsymbol{X}^{\text{con}}$, $\boldsymbol{S}$, D and $\boldsymbol{X}_B^{\text{cat}}$ follows a multinomial logistic regression with main effects for $\boldsymbol{S}$, $\boldsymbol{X}_B^{\text{cat}}$, D and $\boldsymbol{X}^{\text{con}}$. It can also be seen that, because there are no constraints on the parameters of the general location model of expressions (3) and (4) or on $\boldsymbol{\zeta}_1$, $\boldsymbol{\zeta}_2$ and $\boldsymbol{\zeta}_3$, there are also no constraints on the main-effect parameters in this multinomial logistic regression model. Hence, a multinomial logistic regression of a partially observed categorical variable on $\boldsymbol{X}^{\text{con}}$, $\boldsymbol{S}$, D and the elements of $\boldsymbol{X}^{\text{cat}}$ that are not dummy indicators for this categorical variable (with main effects only) constitutes a conditional model that is compatible (in the sense of Definition 1 of Liu et al., 2014) with the general location model of equations (3) and (4).

Web Appendix E: Proof that equations (6)–(8) imply that (2) holds with $\boldsymbol{\beta}_{\text{cat}} = \boldsymbol{\tau} - \boldsymbol{\rho}^\top(\boldsymbol{F} - \boldsymbol{C})\boldsymbol{\xi}$ and $\boldsymbol{\beta}_{\text{con}} = (\boldsymbol{F} - \boldsymbol{C})\boldsymbol{\xi}$

Let $q = \dim(\boldsymbol{X}^{\text{con}})$, and let $\boldsymbol{\xi}_j^* = (\mathbf{0}_{q(j-1)}^\top, \boldsymbol{\xi}^\top, \mathbf{0}_{q(M-j+1)}^\top)^\top$ be a vector of length $(M+1)q$, with $\mathbf{0}_a$ meaning a vector of length a and composed of zeros. From equations (8)–(9) it can be shown that

$$
\begin{aligned}
&\log P(\boldsymbol{X}_1^{\text{con}}, \ldots, \boldsymbol{X}_{M+1}^{\text{con}} \mid \boldsymbol{X}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}}, D_j = 1, D_1 = \ldots = D_{j-1} = D_{j+1} = \ldots = D_{M+1} = 0) \\
&- \log P(\boldsymbol{X}_1^{\text{con}}, \ldots, \boldsymbol{X}_{M+1}^{\text{con}} \mid \boldsymbol{X}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}}, D_1 = 1, D_2 = \ldots = D_{M+1} = 0) \\
&= \begin{bmatrix} \boldsymbol{X}_1^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\rho}\boldsymbol{X}_1^{\text{cat}} - \boldsymbol{\psi}\bar{\boldsymbol{X}}^{\text{cat}} \\ \vdots \\ \boldsymbol{X}_{M+1}^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\rho}\boldsymbol{X}_{M+1}^{\text{cat}} - \boldsymbol{\psi}\bar{\boldsymbol{X}}^{\text{cat}} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\ \boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{\Omega} & \boldsymbol{\Omega} & & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} \\ \boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega} \end{bmatrix}^{-1} \\
&\quad \times (\boldsymbol{\xi}_j^* - \boldsymbol{\xi}_1^*)
\end{aligned} \tag{16}
$$

Now, using the following lemma, it can be shown that

$$
\begin{aligned}
&\begin{bmatrix} \boldsymbol{X}_1^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\rho}\boldsymbol{X}_1^{\text{cat}} - \boldsymbol{\psi}\bar{\boldsymbol{X}}^{\text{cat}} \\ \vdots \\ \boldsymbol{X}_{M+1}^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\rho}\boldsymbol{X}_{M+1}^{\text{cat}} - \boldsymbol{\psi}\bar{\boldsymbol{X}}^{\text{cat}} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\ \boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{\Omega} & \boldsymbol{\Omega} & & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} \\ \boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega} \end{bmatrix}^{-1} \boldsymbol{\xi}_j^* \\
&= (\boldsymbol{X}_j^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\rho}\boldsymbol{X}_j^{\text{cat}} - \boldsymbol{\psi}\bar{\boldsymbol{X}}^{\text{cat}})^\top (\boldsymbol{C} - \boldsymbol{F})\boldsymbol{\xi} \\
&\quad + \sum_{k=1}^{M+1} (\boldsymbol{X}_k^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\rho}\boldsymbol{X}_k^{\text{cat}} - \boldsymbol{\psi}\bar{\boldsymbol{X}}^{\text{cat}})^\top \boldsymbol{F}\boldsymbol{\xi}
\end{aligned} \tag{17}
$$

From equations (6), (8), (16) and (17), it can be shown that

$$
\begin{aligned}
G(\boldsymbol{x}_1^{\text{cat}}, \boldsymbol{x}_1^{\text{con}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}}, \boldsymbol{x}_{M+1}^{\text{con}}) &= \frac{\exp\left\{\begin{array}{l} \boldsymbol{\tau}^\top \boldsymbol{x}_1^{\text{cat}} + \boldsymbol{x}_1^{\text{con}\top}(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi} - \boldsymbol{x}_1^{\text{cat}\top}\boldsymbol{\rho}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi} \\ -\bar{\boldsymbol{x}}_1^{\text{cat}\top}\boldsymbol{\psi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi} \end{array}\right\}}{\sum_{j=1}^{M+1}\exp\left\{\begin{array}{l} \boldsymbol{\tau}^\top \boldsymbol{x}_j^{\text{cat}} + \boldsymbol{x}_j^{\text{con}\top}(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi} - \boldsymbol{x}_j^{\text{cat}\top}\boldsymbol{\rho}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi} \\ -\bar{\boldsymbol{x}}^{\text{cat}\top}\boldsymbol{\psi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi} \end{array}\right\}} \\
&= \frac{\exp\{\boldsymbol{\tau}^\top \boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\xi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{x}_1^{\text{con}} - \boldsymbol{\xi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\rho}\boldsymbol{x}_1^{\text{cat}}\}}{\sum_{j=1}^{M+1}\exp\{\boldsymbol{\tau}^\top \boldsymbol{x}_j^{\text{cat}} + \boldsymbol{\xi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{x}_j^{\text{con}} - \boldsymbol{\xi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\rho}\boldsymbol{x}_j^{\text{cat}}\}} \\
&= \frac{\exp[\{\boldsymbol{\tau} - \boldsymbol{\rho}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi}\}^\top \boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\xi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{x}_1^{\text{con}}]}{\sum_{j=1}^{M+1}\exp[\{\boldsymbol{\tau} - \boldsymbol{\rho}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{\xi}\}^\top \boldsymbol{x}_j^{\text{cat}} + \boldsymbol{\xi}^\top(\boldsymbol{F}-\boldsymbol{C})\boldsymbol{x}_j^{\text{con}}]}
\end{aligned}
$$

Lemma

If $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ are invertible, symmetric $n \times n$ matrices and $M \geqslant 1$, then the inverse of the $(M+1)n \times (M+1)n$ matrix

$$
\begin{bmatrix}
\boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\
\boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\
\vdots & & \ddots & & \vdots \\
\boldsymbol{\Omega} & \boldsymbol{\Omega} & & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} \\
\boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega}
\end{bmatrix}
$$

is given by

$$
\begin{bmatrix}
\boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\
\boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & & \boldsymbol{\Omega} & \boldsymbol{\Omega} \\
\vdots & & \ddots & & \vdots \\
\boldsymbol{\Omega} & \boldsymbol{\Omega} & & \boldsymbol{\Lambda}+\boldsymbol{\Omega} & \boldsymbol{\Omega} \\
\boldsymbol{\Omega} & \boldsymbol{\Omega} & \ldots & \boldsymbol{\Omega} & \boldsymbol{\Lambda}+\boldsymbol{\Omega}
\end{bmatrix}^{-1}
=
\begin{bmatrix}
\boldsymbol{C} & \boldsymbol{F} & \ldots & \boldsymbol{F} & \boldsymbol{F} \\
\boldsymbol{F} & \boldsymbol{C} & & \boldsymbol{F} & \boldsymbol{F} \\
\vdots & & \ddots & & \vdots \\
\boldsymbol{F} & \boldsymbol{F} & & \boldsymbol{C} & \boldsymbol{F} \\
\boldsymbol{F} & \boldsymbol{F} & \ldots & \boldsymbol{F} & \boldsymbol{C}
\end{bmatrix}
$$

where

$$
\begin{aligned}
\boldsymbol{C}^{-1} &= \boldsymbol{\Lambda} + \boldsymbol{\Omega} - \boldsymbol{\Omega}(\boldsymbol{\Lambda} + M\boldsymbol{\Omega})^{-1} M\boldsymbol{\Omega} \\
\boldsymbol{F} &= -(\boldsymbol{\Lambda} + M\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}\boldsymbol{C}
\end{aligned}
$$

This result is easily verified. Note that $\boldsymbol{C}$ and $\boldsymbol{F}$ are symmetric.

Web Appendix F: Proof that linear regression and multinomial logistic regression conditional models are compatible with the general location model of equations (6)–(9)

First, consider a continuous covariate. It follows from expressions (8)–(9) that

$$\boldsymbol{X}_j^{\text{con}} \mid \left\{ \begin{array}{l} \boldsymbol{X}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}}, \boldsymbol{X}_1^{\text{con}}, \ldots, \boldsymbol{X}_{j-1}^{\text{con}}, \boldsymbol{X}_{j+1}^{\text{con}}, \ldots, \boldsymbol{X}_{M+1}^{\text{con}}, \\ D_1 = 1, D_2 = \ldots = D_{M+1} = 0 \end{array} \right\} \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Lambda}') \qquad (18)$$

with

$$\begin{aligned} \boldsymbol{\mu}_j &= \boldsymbol{\eta} + \boldsymbol{\xi} D_j + \boldsymbol{\rho} \boldsymbol{X}_j^{\text{cat}} + \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}} + [\boldsymbol{\Omega}, \ldots, \boldsymbol{\Omega}] \\ &\times \begin{bmatrix} \boldsymbol{C}' & \boldsymbol{F}' & \ldots & \boldsymbol{F}' & \boldsymbol{F}' \\ \boldsymbol{F}' & \boldsymbol{C}' & & \boldsymbol{F}' & \boldsymbol{F}' \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{F}' & \boldsymbol{F}' & & \boldsymbol{C}' & \boldsymbol{F}' \\ \boldsymbol{F}' & \boldsymbol{F}' & \ldots & \boldsymbol{F}' & \boldsymbol{C}' \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\xi} D_1 - \boldsymbol{\rho} \boldsymbol{X}_1^{\text{cat}} - \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}} \\ \vdots \\ \boldsymbol{X}_{j-1}^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\xi} D_{j-1} - \boldsymbol{\rho} \boldsymbol{X}_{j-1}^{\text{cat}} - \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}} \\ \vdots \\ \boldsymbol{X}_{j+1}^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\xi} D_{j+1} - \boldsymbol{\rho} \boldsymbol{X}_{j+1}^{\text{cat}} - \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}} \\ \vdots \\ \boldsymbol{X}_{M+1}^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\xi} D_{M+1} - \boldsymbol{\rho} \boldsymbol{X}_{M+1}^{\text{cat}} - \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}} \end{bmatrix} \end{aligned}$$

where the $M \times M$ matrices $\boldsymbol{C}'$ and $\boldsymbol{F}'$ are the same as the $(M+1) \times (M+1)$ matrices $\boldsymbol{C}$ and $\boldsymbol{F}$ but with M replaced by $(M-1)$ and the final row and final column missing. Hence,

$$
\begin{aligned}
\boldsymbol{\mu}_j &= \boldsymbol{\eta} + \boldsymbol{\xi} D_j + \boldsymbol{\rho} \boldsymbol{X}_j^{\text{cat}} + \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}} \\
&\quad + \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \sum_{k \neq j} (\boldsymbol{X}_k^{\text{con}} - \boldsymbol{\eta} - \boldsymbol{\xi} D_k - \boldsymbol{\rho} \boldsymbol{X}_k^{\text{cat}} - \boldsymbol{\psi} \bar{\boldsymbol{X}}^{\text{cat}}) \\
&= [\boldsymbol{I} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} M] \boldsymbol{\eta} + \boldsymbol{\xi} D_j \\
&\quad + \left[\boldsymbol{\rho} + \frac{1}{M+1} \boldsymbol{\psi} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \frac{M}{M+1} \boldsymbol{\psi} \right] \boldsymbol{X}_j^{\text{cat}} \\
&\quad + \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \sum_{k \neq j} \boldsymbol{X}_k^{\text{con}} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \boldsymbol{\xi} \sum_{k \neq j} D_k \\
&\quad + \left[\frac{1}{M+1} \boldsymbol{\psi} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \boldsymbol{\rho} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \frac{M}{M+1} \boldsymbol{\psi} \right] \sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}} \\
&= [\boldsymbol{I} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} M] \boldsymbol{\eta} + [I(j=1) - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} I(j \neq 1)] \boldsymbol{\xi} \\
&\quad + \left[\boldsymbol{\rho} + \frac{1}{M+1} \boldsymbol{\psi} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \frac{M}{M+1} \boldsymbol{\psi} \right] \boldsymbol{X}_j^{\text{cat}} \\
&\quad + \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \sum_{k \neq j} \boldsymbol{X}_k^{\text{con}} \\
&\quad + \left[\frac{1}{M+1} \boldsymbol{\psi} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \boldsymbol{\rho} - \boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\} \frac{M}{M+1} \boldsymbol{\psi} \right] \\
&\qquad \times \sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}} \qquad (19)
\end{aligned}
$$

Line (19) follows because $D_1, \ldots, D_{M+1}$ sum to one. Also,

$$
\begin{aligned}
\boldsymbol{\Lambda}' &= \boldsymbol{\Lambda} + \boldsymbol{\Omega} - [\boldsymbol{\Omega}, \ldots, \boldsymbol{\Omega}] \begin{bmatrix} \boldsymbol{C}' & \boldsymbol{F}' & \ldots & \boldsymbol{F}' & \boldsymbol{F}' \\ \boldsymbol{F}' & \boldsymbol{C}' & & \boldsymbol{F}' & \boldsymbol{F}' \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{F}' & \boldsymbol{F}' & & \boldsymbol{C}' & \boldsymbol{F}' \\ \boldsymbol{F}' & \boldsymbol{F}' & \ldots & \boldsymbol{F}' & \boldsymbol{C}' \end{bmatrix} \begin{bmatrix} \boldsymbol{\Omega} \\ \vdots \\ \boldsymbol{\Omega} \end{bmatrix} \\
&= \boldsymbol{\Lambda} + \boldsymbol{\Omega} - M\boldsymbol{\Omega}\{\boldsymbol{C}' + (M-1)\boldsymbol{F}'\}\boldsymbol{\Omega}
\end{aligned}
$$

Note that the right-hand side of equation (19) is a linear combination of $\boldsymbol{X}_j^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{con}}$, and that $\boldsymbol{\Lambda}'$ does not depend on any of the variables.

Now partition $\boldsymbol{X}_j^{\text{con}}$ as $\boldsymbol{X}_j^{\text{con}} = (X_{jA}^{\text{con}\top}, \boldsymbol{X}_{jB}^{\text{con}\top})^\top$, where X_{jA}^{con} denotes a single element of

$\boldsymbol{X}_j^{\text{con}}$. It follows from expression (18) that the distribution of X_{jA}^{con} given the other variables is

$$X_{jA}^{\text{con}} \mid \left\{ \begin{array}{l} \boldsymbol{X}_{jB}^{\text{con}}, \boldsymbol{X}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}}, \boldsymbol{X}_1^{\text{con}}, \ldots, \boldsymbol{X}_{j-1}^{\text{con}}, \boldsymbol{X}_{j+1}^{\text{con}}, \ldots, \boldsymbol{X}_{M+1}^{\text{con}}, \\ D_1 = 1, D_2 = \ldots = D_{M+1} = 0 \end{array} \right\} \sim N(\mu_{jA}, \Lambda'')$$

where $\boldsymbol{\mu}_{jA}$ is a linear combination of $\boldsymbol{X}_j^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{con}}$ and $\boldsymbol{X}_{jB}^{\text{con}}$, and where Λ'', like $\boldsymbol{\Lambda}'$, is not a function of the variables.

Since the parameters in equation (8) are unconstrained, it can be seen from expression (18) that the parameters relating the mean of $\boldsymbol{X}_j^{\text{con}}$ to $\boldsymbol{X}_j^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}}$ and $\sum_{k \neq j} \boldsymbol{X}_k^{\text{con}}$ are unconstrained. Thus, the parameters relating the mean of $\boldsymbol{\mu}_{jA}$ to $\boldsymbol{X}_j^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{con}}$ and $\boldsymbol{X}_{jB}^{\text{con}}$ are also unconstrained. Therefore a linear regression of a partially observed element of $\boldsymbol{X}_j^{\text{con}}$ on $\boldsymbol{X}_j^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{con}}$ and the remaining elements of $\boldsymbol{X}_j^{\text{con}}$ (with main effects only and no interactions) constitutes a conditional model that is compatible (in the sense of Definition 1 of Liu et al., 2014) with the general location model of expressions (6)–(8).

Second, consider a categorical covariate. Using equation (9.28) of Schafer (1997), it follows from expressions (6), (7) and (8) that

$$\begin{aligned} &P(\boldsymbol{X}_1^{\text{cat}} = \boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}} = \boldsymbol{x}_{M+1}^{\text{cat}} \mid \boldsymbol{X}_1^{\text{con}}, \ldots, \boldsymbol{X}_{M+1}^{\text{con}}, D_1 = 1, D_2 = \ldots = D_{M+1} = 0) \\ &= \frac{\exp\{b(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}};\ \boldsymbol{\nu}) + \boldsymbol{\tau}^\top \boldsymbol{x}_1^{\text{cat}} + c(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}})\}}{\sum_{\boldsymbol{x}_1^{\text{cat}'}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}'}} \exp\{b(\boldsymbol{x}_1^{\text{cat}'}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}'};\ \boldsymbol{\nu}) + \boldsymbol{\tau}^\top \boldsymbol{x}_1^{\text{cat}'} + c(\boldsymbol{x}_1^{\text{cat}'}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}'})\}} \end{aligned} \tag{20}$$

where

$$
\begin{aligned}
& c(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}}) \\
&= \begin{bmatrix} \boldsymbol{\eta} + \boldsymbol{\xi} D_1 + \boldsymbol{\rho}\boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}} \\ \vdots \\ \boldsymbol{\eta} + \boldsymbol{\xi} D_{M+1} + \boldsymbol{\rho}\boldsymbol{x}_{M+1}^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{C} & \boldsymbol{F} & \ldots & \boldsymbol{F} & \boldsymbol{F} \\ \boldsymbol{F} & \boldsymbol{C} & & \boldsymbol{F} & \boldsymbol{F} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{F} & \boldsymbol{F} & & \boldsymbol{C} & \boldsymbol{F} \\ \boldsymbol{F} & \boldsymbol{F} & \ldots & \boldsymbol{F} & \boldsymbol{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1^{\text{con}} \\ \vdots \\ \boldsymbol{X}_{M+1}^{\text{con}} \end{bmatrix} \\
&\quad - \frac{1}{2} \begin{bmatrix} \boldsymbol{\eta} + \boldsymbol{\xi} D_1 + \boldsymbol{\rho}\boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}} \\ \vdots \\ \boldsymbol{\eta} + \boldsymbol{\xi} D_{M+1} + \boldsymbol{\rho}\boldsymbol{x}_{M+1}^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}} \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{C} & \boldsymbol{F} & \ldots & \boldsymbol{F} & \boldsymbol{F} \\ \boldsymbol{F} & \boldsymbol{C} & & \boldsymbol{F} & \boldsymbol{F} \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{F} & \boldsymbol{F} & & \boldsymbol{C} & \boldsymbol{F} \\ \boldsymbol{F} & \boldsymbol{F} & \ldots & \boldsymbol{F} & \boldsymbol{C} \end{bmatrix} \\
&\qquad \times \begin{bmatrix} \boldsymbol{\eta} + \boldsymbol{\xi} D_1 + \boldsymbol{\rho}\boldsymbol{x}_1^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}} \\ \vdots \\ \boldsymbol{\eta} + \boldsymbol{\xi} D_{M+1} + \boldsymbol{\rho}\boldsymbol{x}_{M+1}^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}} \end{bmatrix} \\
&= \sum_{j=1}^{M+1} (\boldsymbol{\eta} + \boldsymbol{\xi} D_j + \boldsymbol{\rho}\boldsymbol{x}_j^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}})^\top \left\{ \boldsymbol{C} \left(\boldsymbol{X}_j^{\text{con}} - \frac{1}{2}(\boldsymbol{\eta} + \boldsymbol{\xi} D_j + \boldsymbol{\rho}\boldsymbol{x}_j^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}}) \right) \right. \\
&\qquad \left. + \boldsymbol{F} \sum_{k \neq j} \left(\boldsymbol{X}_k^{\text{con}} - \frac{1}{2}(\boldsymbol{\eta} + \boldsymbol{\xi} D_k + \boldsymbol{\rho}\boldsymbol{x}_k^{\text{cat}} + \boldsymbol{\psi}\bar{\boldsymbol{x}}^{\text{cat}}) \right) \right\} \qquad (21)
\end{aligned}
$$

Now partition $\boldsymbol{X}_j^{\text{cat}}$ as $\boldsymbol{X}_j^{\text{cat}} = (\boldsymbol{X}_{jA}^{\text{cat}\top}, \boldsymbol{X}_{jB}^{\text{cat}\top})^\top$, where $\boldsymbol{X}_{jA}^{\text{cat}}$ denotes the subvector of $\boldsymbol{X}_j^{\text{cat}}$ that consists of the dummy indicators for a single categorical covariate, and partition $\boldsymbol{x}_j^{\text{cat}}$ analogously. To obtain the distribution of $\boldsymbol{X}_{jA}^{\text{cat}}$ given $\boldsymbol{X}_{jB}^{\text{cat}}, \boldsymbol{X}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{j-1}^{\text{cat}}, \boldsymbol{X}_{j+1}^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}}$, $\boldsymbol{X}_1^{\text{con}}, \ldots, \boldsymbol{X}_{M+1}^{\text{con}}$, $D_1 = 1$ and $D_2 = \ldots = D_{M+1} = 0$, we can take equation (20) and ignore

terms not involving $\boldsymbol{x}_{jA}^{\text{cat}}$. Using equation (21) it can be shown that

$$\begin{aligned} & b(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}};\ \boldsymbol{\nu}) + \boldsymbol{\tau}^\top \boldsymbol{x}_1^{\text{cat}} + c(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}}) \\ & = \boldsymbol{u}_{1j}^\top \boldsymbol{x}_{jA}^{\text{cat}} + \boldsymbol{x}_{jB}^{\text{cat}\top} \boldsymbol{U}_2 \boldsymbol{x}_{jA}^{\text{cat}} + \boldsymbol{X}_j^{\text{con}\top} \boldsymbol{U}_3 \boldsymbol{x}_{jA}^{\text{cat}} + \sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}\top} \boldsymbol{U}_4 \boldsymbol{x}_{jA}^{\text{cat}} + \sum_{k \neq j} \boldsymbol{X}_k^{\text{con}\top} \boldsymbol{U}_5 \boldsymbol{x}_{jA}^{\text{cat}} \\ & \quad + \text{terms not involving } \boldsymbol{x}_{jA}^{\text{cat}} \end{aligned} \tag{22}$$

where $\boldsymbol{u}_{1j}$ is a vector function of $\boldsymbol{\nu}$ and the parameters in equation (21), and $\boldsymbol{U}_2, \ldots, \boldsymbol{U}_5$ are matrix functions of the same set of parameters. From equations (20) and (22) it can be seen that the distribution of $\boldsymbol{X}_{jA}^{\text{cat}}$ given $\boldsymbol{X}_{jB}^{\text{cat}}$, $\boldsymbol{X}_1^{\text{cat}}, \ldots, \boldsymbol{X}_{j-1}^{\text{cat}}, \boldsymbol{X}_{j+1}^{\text{cat}}, \ldots, \boldsymbol{X}_{M+1}^{\text{cat}}$, $\boldsymbol{X}_1^{\text{con}}, \ldots, \boldsymbol{X}_{M+1}^{\text{con}}$, $D_1 = 1$ and $D_2 = \ldots = D_{M+1} = 0$ has the form of a multinomial logistic regression with main effects for $\boldsymbol{X}_j^{\text{con}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{con}}$, $\sum_{k \neq j} \boldsymbol{X}_k^{\text{cat}}$ and $\boldsymbol{X}_{jB}^{\text{cat}}$.

The inclusion of $b(\boldsymbol{x}_1^{\text{cat}}, \ldots, \boldsymbol{x}_{M+1}^{\text{cat}};\ \boldsymbol{\nu})$ with unconstrained $\boldsymbol{\nu}$ ensures that $\boldsymbol{u}_{1j}$, $\boldsymbol{U}_2$ and $\boldsymbol{U}_4$ are unconstrained. Therefore, it only remains to check that $\boldsymbol{U}_3$ and $\boldsymbol{U}_5$ are unconstrained. It can be shown that

$$\begin{aligned} \boldsymbol{U}_3 &= \boldsymbol{C}\{\boldsymbol{\rho}_A + \boldsymbol{\psi}_A (M+1)^{-1}\} + \boldsymbol{F}\boldsymbol{\psi}_A M (M+1)^{-2} \\ \boldsymbol{U}_5 &= \boldsymbol{C}\boldsymbol{\psi}_A (M+1)^{-1} + \boldsymbol{F}\{\boldsymbol{\rho}_A + \boldsymbol{\psi}_A M (M+1)^{-1}\} \end{aligned}$$

where $\boldsymbol{\rho}_A$ and $\boldsymbol{\psi}_A$ are given by the partitions $\boldsymbol{\rho} = [\boldsymbol{\rho}_A \ \ \boldsymbol{\rho}_B]$ and $\boldsymbol{\psi} = [\boldsymbol{\psi}_A \ \ \boldsymbol{\psi}_B]$. Since $\boldsymbol{\rho}_A$ and $\boldsymbol{\psi}_A$ are unconstrained, so are $\boldsymbol{U}_3$ and $\boldsymbol{U}_5$. So, none of the parameters in the multinomial logistic regression is constrained. Hence, this multinomial logistic regression is compatible (in the sense of Definition 1 of Liu et al., 2014) with the general location model of expressions (6)–(8).

Web Appendix G: Additional simulation studies with misspecified models

We simulated datasets using data-generating mechanisms that made the imputation models of all our MI methods misspecified. In the first, data were simulated with an interaction

between S^{cat} and S^{con}; in the second, X^{conA} and X^{conB} were log-normally distributed. We now detail these two data-generating mechanisms, before giving the results.

Sensitivity analysis 1: interaction between S^{cat} and S^{con}

The data-generating process used in the main simulations (Section 6) assumed that the conditional distribution of $(X^{\text{conA}}, X^{\text{conB}})$ given $X^{\text{cat}}, S^{\text{cat}}, S^{\text{con}}$ and D was bivariate normal with univariate marginal distributions $N(0.5X^{\text{cat}}+0.5S^{\text{cat}}+0.5S^{\text{con}}+0.5D, 1)$ and covariance 0.5. We modified this data-generating process by changing the univariate marginal distributions to $N(0.5X^{\text{cat}}+0.5S^{\text{cat}}+0.5S^{\text{con}}+0.5S^{\text{cat}}S^{\text{con}}+0.5D, 1)$, thus including an interaction term that was not present in the imputation models. The rest of the data-generating process was the same as that used in the main simulations, including the application of the MCAR, MAR-A and MAR-B missingness mechanisms with 10% or 25% missingness. One thousand simulated datasets, each with $N = 500$ cases and 500 matched controls ($M = 1$), were generated.

Sensitivity analysis 2: log normally distributed X^{conA} and X^{conB}

We simulated data for a matched case-control study as follows. First, data on a population of 15000 independent individuals were generated. For each individual in this population, variables S^{cat}, S^{con}, X^{cat}, X^{conA}, X^{conB} and D were generated. Binary variable S^{cat} and continuous variable S^{con} were generated independently using $P(S^{\text{cat}} = 1) = 0.5$ and $S^{\text{con}} \sim \text{Normal}(0, 1)$. Binary variable X^{cat} was generated with logit $P(X^{\text{cat}} = 1 \mid S^{\text{cat}}, S^{\text{con}}) = -2.5+0.5S^{\text{cat}}+0.5S^{\text{con}}$. Continuous variables X^{conA} and X^{conB} were generated using $X^{\text{conA}} = 0.5X^{\text{cat}} + 0.5S^{\text{cat}} + 0.5S^{\text{con}} + \exp(e_A)$ and $X^{\text{conB}} = 0.5X^{\text{cat}} + 0.5S^{\text{cat}} + 0.5S^{\text{con}} + \exp(e_B)$, where (e_A, e_B) was bivariate normally distributed with zero mean and unit variance, independently of S^{cat}, S^{con} and X^{cat}. The covariance of (e_A, e_B) was chosen so that the covariance between X^{conA} and X^{conB} conditional on X^{cat}, S^{cat} and S^{con} was approximately 0.5, as

it was in the main simulations. Binary variable D was generated using logit $P(D = 1 \mid S^{\text{cat}}, S^{\text{con}}, X^{\text{cat}}, X^{\text{conA}}, X^{\text{conB}}) = -4.68 + 0.25S^{\text{cat}} + 0.25S^{\text{con}} + \frac{5}{12}X^{\text{cat}} + \frac{1}{3}X^{\text{conA}} + \frac{1}{3}X^{\text{conB}}$. The intercept of -4.68 was chosen to give $P(D = 1) = 0.05$, i.e. a 5% population prevalence of disease.

Having generated data on this population of 15000 individuals, 500 individuals with $D = 1$ were randomly drawn from it. These are the 500 cases. One individual with $D = 0$ was then matched with each case on S^{cat} and S^{con}. These are the 500 matching controls. Finally, missingness was randomly imposed on this matched case-control sample using the same MCAR, MAR-A and MAR-B mechanisms with 10% or 25% missingness that were used in the main simulation study.

This entire process (i.e. simulation of population and sampling of cases and controls) was repeated 1000 times to generate 1000 matched case-control study datasets.

Results

Web Tables 11–13 and 14–16 show the results for the first and second sensitivity analyses, respectively. It can be seen that the MI methods are reasonably robust to the imputation model misspecifications.

Web Appendix H: Application of MI methods in Stata and R

Here we provide the Stata or R code that we used to perform MI in the simulation study when there were $M = 4$ controls per case, for each of the methods outlined in the paper, along with a specimen simulated dataset. Text files of the Stata and R code given below and the files containing the simulated dataset are available as a web supplement at the Biometrics website. The dataset is provided as a Stata file (simulated_data.dta) and as a comma-separated-values file (simulated_data.csv).

In the dataset the three covariates X^{cat}, X^{conA} and X^{conB} are denoted `xcat`, `xconA` and `xconB`, respectively. The categorical and continuous matching variables S^{cat} and S^{con} are denoted respectively as `scat` and `scon`, and the case or control status is denoted `d`. There are missing values in X^{conA} and X^{conB}, which were generated completely at random to give approximately 25% of the values missing in each variable.

We specify a seed at the start of each block of code, so that the results are reproducible. The results of applying the code to the simulated dataset provided are shown in Web Table 17. The results from the complete case analysis are also shown; this analysis can be performed in, for example, Stata using `clogit d xcat xconA xconB , group(set)`.

FCS MI using matching variables

This analysis was performed in Stata using the `ice` command, as shown below.

```
use simulated_data.dta, clear

set seed 780423

ice d xcat xconA xconB i.scat scon, /*
*/ eq(xcat: d xconA xconB _Iscat_1 scon, xconA: d xcat xconB _Iscat_1 scon)/*
*/ m(50) clear

* fit the analysis model to each imputed dataset and combine using Rubin's Rules
mim: clogit d xcat xconA xconB , group(set)
```

Normal MI using matching variables

This analysis was performed in Stata using `mi impute`, as shown below. Adaptive rounding is used post-imputation to assign values 0 or 1 for missing values in `xcat`.

```
use simulated_data.dta, clear

set seed 780423
```

```
mi set mlong
mi register imputed xcat xconA
mi register regular d xconB scat scon

mi impute mvn xcat xconA = d xconB i.scat scon, add(50)

* apply adaptive rounding for xcat
mi passive: egen thres_mean=mean(xcat)
mi passive: gen thres=thres_mean-invnormal(thres_mean)*sqrt(thres_mean*(1-thres_mean))
replace xcat=0 if xcat<=thres & _mi_m!=0 & xcat!=0 & xcat!=1
replace xcat=1 if xcat>thres & _mi_m!=0 & xcat!=0 & xcat!=1

* fit the analysis model to each imputed dataset and combine using Rubin's Rules
mim: clogit d xcat xconA xconB , group(set)
```

Adaptive rounding does not apply for non-binary categorical variables, which should be represented by a series of dummy variables. Non-binary categorical variables can be handled using the approach outlined by Carpenter and Kenward (2013, Section 5.2). In this approach, if the imputed values of all the dummy variables are less than 0.5, then the most common category is assigned; otherwise, the category with the highest imputed value is assigned. The categories of the categorical variable are first ordered so that the most frequently occurring category is the reference category. This was the approach used in the EPIC example.

Latent normal MI using matching variables

This analysis was performed in R using the `jomo1mix` function in the 'jomo' package. The 'jomo', 'survival' and 'mice' libraries are required. The `jomo1mix` function requires a dataframe containing continuous variables with missing values (`data.cont`), a dataframe containing categorical variables with missing values (`data.cat`), and a dataframe containing the outcome, the fully observed variables (including the matching variables) and a vector of ones (`data.x.matrix`). The number of categories in the categorical variable with missing

values is indicated by `Y_numcat=2` (this becomes a vector if there is more than one categorical variable with missing values). Categorical variables with missing values must take ordered integer values starting at 1. A loop is used to fit the conditional logistic regression in each imputed dataset, and the stored results are combined using the `pool.scalar` function, which is in the 'mice' package. In the output from `pool.scalar` the pooled estimate is given by `qbar` and its variance is given by `t`.

```
data<-read.table("simulated_data.csv",sep=",",header=T)

set.seed(780423)

n.imp <- 50 # number of imputations

data.cont<-data.frame(data$xconA)
data.cat<-data.frame(data$xcat)+1
data.x.matrix<-data.frame(rep(1,length(data$ccid)),data$scat,data$scon,data$xconB,data$d)

myjomo<-jomo1mix(Y_con=data.cont, Y_cat=data.cat, Y_numcat=2, X=data.x.matrix,nimp=n.imp)

#fit the analysis model to each imputed dataset
coef<-matrix(nrow=n.imp,ncol=3)
var<-matrix(nrow=n.imp,ncol=3)
for(k in 1:n.imp){
  imp.data<-data.frame(myjomo[myjomo$Imputation==k,],data$set)
  model.imp<-clogit(data.d~data.xcat+data.xconA+data.xconB+strata(data.set),data=imp.data)
  coef[k,]<-model.imp$coef
  var[k,]<-diag(model.imp$var)
}

# combine using Rubin's Rules
pool.scalar(coef[,1],var[,1])
pool.scalar(coef[,2],var[,2])
pool.scalar(coef[,3],var[,3])
```

FCS MI using matched set

This analysis was performed in Stata using the `ice` command, as shown below. First, the dataset was transformed to 'wide' format using the reshape command, i.e. so that there is one row per matched set. Then, within each matched set, $\sum_{k \neq j} X_k^{\texttt{cat}}$, $\sum_{k \neq j} X_k^{\texttt{conA}}$ and $\sum_{k \neq j} X_k^{\texttt{conB}}$ were calculated for $j = 1, \ldots, 5$. These are denoted as 'xcatsum1', ..., 'xcatsum5', 'xconAsum1', ..., 'xconAsum5', and 'xconBsum1', ..., 'xconBsum5'. Note the following features of the ice command used for this analysis. First, the option 'eq(xcat1: xconA1 xconB1 xcatsum1 xconAsum1 xconBsum1' tells ice that the conditional model for xcat1 is a regression of xcat1 on xconA1, xconB1, xcatsum1, xconAsum1 and xconBsum1. Without this option, xcat1 would be regressed (by default) on all the other covariates. Second, the option 'passive(xcatsum1:(xcat2+xcat3+xcat4+xcat5)' tells ice what xcatsum1 means: it is a variable that is equal to the sum of xcat2, xcat3, xcat4 and xcat5.

Finally, the imputed datasets were transformed back to 'long' format, i.e. so that there is one row per individual (we used the 'reshape' command in Stata), before analysing them using conditional logistic regression and combining the results using Rubin's Rules.

```
use simulated_data.dta, clear

set seed 780423

reshape wide

forval i=1/5{
gen xcatsum‘i’=.
gen xconAsum‘i’=.
}

gen xconBsum1=xconB2+xconB3+xconB4+xconB5
gen xconBsum2=xconB1+xconB3+xconB4+xconB5
gen xconBsum3=xconB1+xconB2+xconB4+xconB5
```

```
gen xconBsum4=xconB1+xconB2+xconB3+xconB5
gen xconBsum5=xconB1+xconB2+xconB3+xconB4

ice d1 xcat1 xconA1 xconB1 d2 xcat2 xconA2 xconB2 d3 xcat3 xconA3 xconB3 d4 xcat4 xconA4 xconB4 /*
*/ d5 xcat5 xconA5 xconB5 xcatsum1 xcatsum2 xcatsum3 xcatsum4 xcatsum5 xconAsum1 xconAsum2 /*
*/ xconAsum3 xconAsum4 xconAsum5 xconBsum1 xconBsum2 xconBsum3 xconBsum4 xconBsum5, /*
*/ eq(xcat1: xconA1 xconB1 xcatsum1 xconAsum1 xconBsum1,/*
*/   xcat2: xconA2 xconB2 xcatsum2 xconAsum2 xconBsum2,/*
*/   xcat3: xconA3 xconB3 xcatsum3 xconAsum3 xconBsum3,/*
*/   xcat4: xconA4 xconB4 xcatsum4 xconAsum4 xconBsum4,/*
*/   xcat5: xconA5 xconB5 xcatsum5 xconAsum5 xconBsum5,/*
*/   xconA1: xcat1 xconB1 xconAsum1 xcatsum1 xconBsum1,/*
*/   xconA2: xcat2 xconB2 xconAsum2 xcatsum2 xconBsum2,/*
*/   xconA3: xcat3 xconB3 xconAsum3 xcatsum3 xconBsum3,/*
*/   xconA4: xcat4 xconB4 xconAsum4 xcatsum4 xconBsum4,/*
*/   xconA5: xcat5 xconB5 xconAsum5 xcatsum5 xconBsum5) /*
*/ passive(xcatsum1:(xcat2+xcat3+xcat4+xcat5)\/*
*/         xcatsum2:(xcat1+xcat3+xcat4+xcat5)\/*
*/         xcatsum3:(xcat2+xcat1+xcat4+xcat5)\/*
*/         xcatsum4:(xcat2+xcat3+xcat1+xcat5)\/*
*/         xcatsum5:(xcat2+xcat3+xcat4+xcat1)\/*
*/         xconAsum1:(xconA2+xconA3+xconA4+xconA5)\/*
*/         xconAsum2:(xconA1+xconA3+xconA4+xconA5)\/*
*/         xconAsum3:(xconA2+xconA1+xconA4+xconA5)\/*
*/         xconAsum4:(xconA2+xconA3+xconA1+xconA5)\/*
*/         xconAsum5:(xconA2+xconA3+xconA4+xconA1)/*
*/        ) m(50) saving(name_of_output_file.dta, replace) clear

use name_of_output_file.dta, clear

mi import ice, automatic

mi reshape long d xcat xconA xconB, i(set) j(setpos)

* fit the analysis model to each imputed dataset and combine using Rubin's Rules
mim: clogit d xcat xconA xconB , group(set)
```

Normal MI using matched set

This analysis was performed in R using the `pan` function in the 'pan' package. The 'pan', 'survival' and 'mice' libraries are required. The `pan` function requires a data frame containing the model covariates (`y`), a data frame containing the outcome and a vector of 1s (`x`), and a data frame containing the matched set variable (`set`). The model covariates without missing values could alternatively be included in `x`. It is necessary to set up an array that stores the imputed datasets (`y.imp`) and also to specify the random seeds used to perform the imputations (`seeds.n.imp`). Adaptive rounding is used post-imputation to assign values 0 or 1 for missing values in `xcat` (note that a different approach is required for non-binary categorical variables, as described in the section on Normal MI using matching variables). A loop is used to fit the conditional logistic regression to each imputed dataset and the stored results are combined using the `pool.scalar` function, which is in the 'mice' package. In the output from `pool.scalar` the pooled estimate is given by `qbar` and its variance is given by `t`.

```
data<-read.table("simulated_data.csv",sep=",",header=T)

set.seed(780423)

N <- 500 # number of matching sets
setsize <- 5 # the number of individuals in a matching set
n.imp <- 50 # number of imputations
nvar <- 3 # number of covariates (xcat, xconA and xconB in this case)
set.seed(10)
seeds <- floor( runif(1000) * 1e6 )

y <- as.matrix( data[, c("xcat", "xconA", "xconB")] )
set <- data[, "set"]
x <- cbind(1, data[, "d"])

y.imp <- array(0, c(N*setsize, nvar, n.imp)) # Set up array to store imputations
```

```
# Get random seeds for all imputations

seeds.n.imp <- floor( runif(n.imp) * 1e6 )

# The first imputation is performed separately, then subsequent imputations
# restart Gibbs sampler from the final state of a previous run

imp.pan <- pan(y=y, subj=set, pred=x, xcol=1:2, zcol=1,
               prior=list(a=nvar, Binv=diag(nvar, nvar), c=nvar, Dinv=diag(nvar, nvar)),
               seeds.n.imp[1], iter=1000)

# first imputed dataset
y.imp[,,1] <- imp.pan$y
# apply adaptive rounding
thres.mean<-mean(y.imp[,1,1])
thres<-thres.mean-qnorm(thres.mean)*sqrt(thres.mean*(1-thres.mean))
assign_zero<-(y.imp[,1,1]<=thres & y.imp[,1,1]!=0 & y.imp[,1,1]!=1)|y.imp[,1,1]==0
y.imp[,1,1]<-ifelse(assign_zero==1,0,1)

# imputations 2:n.imp
for (j in 2:n.imp)
{
 imp.pan <- pan(y=y, subj=set, pred=x, xcol=1:2, zcol=1,
                prior=list(a=nvar, Binv=diag(nvar, nvar), c=nvar, Dinv=diag(nvar, nvar)),
                seed=seeds.n.imp[j], iter=1000, start=imp.pan$last)
 y.imp[,,j] <- imp.pan$y
 # apply adaptive rounding
 thres.mean<-mean(y.imp[,1,j])
 thres<-thres.mean-qnorm(thres.mean)*sqrt(thres.mean*(1-thres.mean))
 assign_zero<-(y.imp[,1,j]<=thres & y.imp[,1,j]!=0 & y.imp[,1,j]!=1)|y.imp[,1,j]==0
 y.imp[,1,j]<-ifelse(assign_zero==1,0,1)
}

# fit the analysis model to each imputed dataset
coef<-matrix(nrow=n.imp,ncol=3)
var<-matrix(nrow=n.imp,ncol=3)
```

```
for (j in 1:n.imp){
  model.imp<-clogit(data$d ~ y.imp[,,j] + strata(data$set))
  coef[j,]<-model.imp$coef
  var[j,]<-diag(model.imp$var)
}

# combine using Rubin's Rules
pool.scalar(coef[,1],var[,1])
pool.scalar(coef[,2],var[,2])
pool.scalar(coef[,3],var[,3])
```

Latent normal MI using matched set

This analysis was performed in R using the `jomo1ranmix` function in the 'jomo' package. The 'jomo', 'survival' and 'mice' libraries are required. Like `jomo1mix`, the `jomo1ranmix` function requires a dataframe containing continuous variables with missing values (`data.cont`), a dataframe containing categorical variables with missing values (`data.cat`), and a dataframe containing the outcome, the fully observed variables (including the matching variables) and a vector of ones (`data.x.matrix`). The number of categories in the categorical variable with missing values is indicated by `Y_numcat=2`. The clustering variable, which here is matched set, is indicated in `jomo1ranmix` by `clus=set`. The clustering variable must take ordered integer values starting at 0. As with `jomo1mix`, categorical variables with missing values must take ordered integer values starting at 1. A loop is used to fit the conditional logistic regression to each imputed dataset and the stored results are combined using the `pool.scalar` function, which is in the 'mice' package. In the output from `pool.scalar` the pooled estimate is given by `qbar` and its variance is given by `t`.

```
data<-read.table("simulated_data.csv",sep=",",header=T)

set.seed(780423)

set<-data.frame(data$set-1)
```

```
n.imp <- 50 # number of imputations

data.cont<-data.frame(data$xconA)
data.cat<-data.frame(data$xcat)+1
data.x.matrix<-data.frame(rep(1,length(data$ccid)),data$xconB,data$d)

myjomo<-jomo1ranmix(Y_con=data.cont, Y_cat=data.cat, Y_numcat=2, X=data.x.matrix,
                    clus=set,nimp=n.imp)

coef<-matrix(nrow=n.imp,ncol=3)
var<-matrix(nrow=n.imp,ncol=3)
for(k in 1:n.imp){
  imp.data<-data.frame(myjomo[myjomo$Imputation==k,],data$set)
  model.imp<-clogit(data.d~data.xcat+data.xconA+data.xconB+strata(data.set),data=imp.data)
  coef[k,]<-model.imp$coef
  var[k,]<-diag(model.imp$var)
}

# combine using Rubin's Rules
pool.scalar(coef[,1],var[,1])
pool.scalar(coef[,2],var[,2])
pool.scalar(coef[,3],var[,3])
```

References

Carpenter J. and Kenward M. (2013) *Multiple Imputation and its Applications*. Wiley, New Jersey.

Hughes, R.A., White, I.R., Seaman, S.R., Carpenter, J.R., Tilling, K. and Sterne, J.A.C. (2014) Joint modelling rationale for chained equations. *BMC Medical Research Methodology* 14, 28.

Lee, K.J. and Carlin, J.B. (2010) Multiple imputation for missing data: fully conditional

specification versus multivariate normal imputation. *American Journal of Epidemiology* 171, 624–632.

Little, R.J.A. and Rubin, D.B. (2002) *Statistical Analysis With Missing Data*. Wiley, New Jersey.

Liu, J., Gelman, A., Hill, J., Su, Y.S. and Kropko, J. (2014) On the stationary distribution of iterative imputations. *Biometrika* 101, 155–173.

Schafer J.L. (1997) *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.

van Buuren, S. (2012) *Flexible Imputation of Missing Data*. Chapman & Hall/CRC.

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

[Table 14 about here.]

[Table 15 about here.]

[Table 16 about here.]

[Table 17 about here.]

[Table 18 about here.]

[Table 19 about here.]

[Table 20 about here.]

[Table 21 about here.]

[Table 22 about here.]

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
10% missing															
Complete cases	0.431	0.264	0.256	70.0	96	0.337	0.100	0.102	10.1	96	0.340	0.103	0.102	10.6	96
Match var: FCS	0.429	0.224	0.219	50.2	94	0.335	0.083	0.087	6.92	96	0.338	0.085	0.084	7.17	94
Normal	0.410	0.214	0.218	45.7	96	0.336	0.083	0.087	6.85	96	0.339	0.084	0.083	7.16	95
Latent norm	0.435	0.223	0.218	50.0	95	0.330	0.081	0.087	6.63	96	0.340	0.084	0.084	7.12	94
Match set: FCS	0.429	0.225	0.219	51.0	95	0.334	0.084	0.087	7.05	96	0.338	0.085	0.084	7.23	94
Normal	0.420	0.221	0.223	49.0	96	0.340	0.086	0.089	7.44	96	0.335	0.086	0.084	7.34	95
Latent norm	0.437	0.226	0.220	51.6	95	0.320	0.082	0.087	6.83	96	0.340	0.085	0.084	7.35	95
25% missing															
Complete cases	0.449	0.379	0.377	145	96	0.341	0.144	0.149	20.8	97	0.342	0.150	0.149	22.4	96
Match var: FCS	0.431	0.240	0.241	57.6	96	0.336	0.090	0.096	8.06	96	0.337	0.087	0.086	7.51	95
Normal	0.386	0.215	0.235	47.1	97	0.338	0.090	0.095	8.04	97	0.341	0.086	0.086	7.49	95
Latent norm	0.446	0.238	0.241	57.7	96	0.322	0.085	0.095	7.31	97	0.343	0.085	0.086	7.39	95
Match set: FCS	0.430	0.247	0.243	61.0	95	0.335	0.094	0.097	8.81	96	0.339	0.088	0.086	7.69	95
Normal	0.407	0.238	0.251	56.8	96	0.350	0.098	0.101	9.93	96	0.329	0.090	0.088	8.08	94
Latent norm	0.455	0.249	0.247	63.7	95	0.300	0.085	0.095	8.27	95	0.344	0.088	0.088	7.89	95

Table 1: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MCAR missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.418	0.150	0.144	22.4	94	0.334	0.058	0.061	3.37	97	0.337	0.062	0.061	3.89	94
10% missing															
Complete cases	0.419	0.179	0.169	32.1	94	0.333	0.069	0.071	4.72	96	0.338	0.073	0.071	5.38	94
Match var: FCS	0.418	0.158	0.153	25.1	95	0.333	0.062	0.065	3.82	97	0.337	0.064	0.062	4.07	94
Normal	0.407	0.154	0.152	23.8	96	0.335	0.062	0.064	3.82	96	0.339	0.064	0.062	4.06	94
Latent norm	0.424	0.158	0.153	25.1	94	0.329	0.061	0.065	3.69	97	0.339	0.063	0.062	4.05	94
Match set: FCS	0.415	0.159	0.153	25.4	94	0.332	0.062	0.065	3.80	96	0.338	0.064	0.062	4.08	94
Normal	0.411	0.157	0.154	24.7	95	0.336	0.062	0.065	3.89	97	0.337	0.064	0.062	4.08	94
Latent norm	0.424	0.159	0.153	25.2	95	0.320	0.060	0.065	3.83	97	0.340	0.064	0.062	4.15	94
25% missing															
Complete cases	0.411	0.242	0.225	58.5	93	0.337	0.089	0.093	7.90	97	0.338	0.095	0.093	8.99	95
Match var: FCS	0.417	0.176	0.168	30.8	94	0.335	0.069	0.071	4.75	97	0.337	0.066	0.064	4.36	94
Normal	0.389	0.164	0.165	27.7	95	0.338	0.069	0.071	4.73	96	0.340	0.066	0.064	4.37	94
Latent norm	0.432	0.174	0.168	30.4	94	0.323	0.066	0.070	4.42	97	0.343	0.065	0.064	4.37	94
Match set: FCS	0.409	0.176	0.168	31.0	94	0.333	0.068	0.071	4.68	96	0.338	0.066	0.064	4.41	94
Normal	0.398	0.170	0.170	29.3	95	0.343	0.070	0.072	5.04	95	0.335	0.067	0.064	4.47	94
Latent norm	0.431	0.177	0.170	31.4	94	0.302	0.065	0.070	5.23	95	0.344	0.067	0.065	4.59	94

Table 2: Results from 1000 simulated datasets of $N = 500$ cases and $M = 4$ controls per case with MCAR missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	β_{cat}				β_{conA}			
	bias	ratio SE	ratio estSE	rel. eff.	bias	ratio SE	ratio estSE	rel. eff.
M=1								
Complete data	0.010	1.000	1.000	100.0	0.003	1.000	1.000	100.0
Match var: FCS	0.014	1.136	1.188	77.4	0.002	1.142	1.161	76.7
Normal	−0.046	1.018	1.156	92.2	0.006	1.137	1.157	77.2
Latent norm	0.026	1.126	1.183	78.1	−0.010	1.081	1.150	84.4
Match set: FCS	0.013	1.181	1.207	71.7	−0.002	1.196	1.184	70.0
Normal	−0.015	1.132	1.245	77.9	0.014	1.251	1.223	62.7
Latent norm	0.035	1.187	1.220	69.7	−0.035	1.082	1.157	73.0
M=4								
Complete data	0.001	1.000	1.000	100.0	0.001	1.000	1.000	100.0
Match var: FCS	−0.001	1.195	1.200	70.0	0.002	1.188	1.172	70.8
Normal	−0.051	1.108	1.171	73.9	0.005	1.184	1.169	70.9
Latent norm	0.010	1.177	1.195	72.0	−0.010	1.129	1.163	76.6
Matchset: FCS	−0.006	1.220	1.210	67.3	−0.002	1.190	1.187	70.5
Normal	−0.036	1.149	1.206	72.0	0.007	1.215	1.196	67.0
Latent norm	0.008	1.200	1.211	69.3	−0.034	1.126	1.168	62.4

Table 3: Biases ('bias'), ratios of empirical SEs ('ratio SE'), ratios of mean estimated SEs ('ratio empSE'), and relative efficiencies (%, 'rel. eff.') of six MI methods when $N = 500$ and $p_{\text{miss}} = 0.25$. Ratios and relative efficiencies are calculated relative to the corresponding complete-data estimators. Each reported ratio or relative efficiency is the average over three ratios or relative efficiencies: one from each of the MCAR, MAR-A and MAR-B scenarios. Reported biases are the signed average absolute bias over these three scenarios.

	β_{cat}				β_{conA}			
	bias	ratio SE	ratio estSE	rel. eff.	bias	ratio SE	ratio estSE	rel. eff.
M=1								
Complete data	0.038	1.000	1.000	100.0	0.018	1.000	1.000	100.0
Match var: FCS	0.048	1.199	1.227	69.7	0.025	1.164	1.192	73.5
Normal	−0.017	1.068	1.186	88.3	0.027	1.150	1.185	75.2
Latent norm	0.066	1.193	1.219	69.9	0.008	1.080	1.178	86.4
Match set: FCS	0.041	1.221	1.277	67.3	0.022	1.234	1.231	65.7
Normal	0.029	1.160	1.309	74.7	0.051	1.321	1.295	55.6
Latent norm	0.099	1.304	1.301	57.9	−0.013	1.085	1.203	85.3
M=4								
Complete data	−0.015	1.000	1.000	100.0	0.003	1.000	1.000	100.0
Match var: FCS	−0.018	1.200	1.217	69.5	0.005	1.130	1.183	78.3
Normal	−0.065	1.099	1.184	80.3	0.008	1.124	1.176	78.9
Latent norm	−0.009	1.191	1.210	70.7	−0.008	1.073	1.172	86.6
Match set: FCS	−0.037	1.157	1.213	74.3	−0.011	1.091	1.183	83.7
Normal	−0.040	1.136	1.228	76.7	0.012	1.172	1.211	72.5
Latent norm	−0.010	1.234	1.237	65.8	−0.026	1.079	1.185	83.6

Table 4: Biases ('bias'), ratios of empirical SEs ('ratio SE'), ratios of mean estimated SEs ('ratio empSE'), and relative efficiencies (%, 'rel. eff.') of six MI methods when $N = 100$ and $p_{\text{miss}} = 0.25$. Ratios and relative efficiencies are calculated relative to the corresponding complete-data estimators. Each reported ratio or relative efficiency is the average over three ratios or relative efficiencies: one from each of the MCAR, MAR-A and MAR-B scenarios. Reported biases are the signed average absolute bias over these three scenarios.

Method	log OR	SE	95% CI	p-value
Complete cases	−0.196	0.126	(−0.444, 0.052)	0.121
MI using matching variables:				
FCS	−0.176	0.104	(−0.380, 0.027)	0.090
Normal	−0.177	0.104	(−0.380, 0.027)	0.088
Latent normal	−0.176	0.104	(−0.380, 0.027)	0.089
MI using matched set:				
FCS	−0.175	0.104	(−0.378, 0.028)	0.092
Normal	−0.181	0.104	(−0.384, 0.023)	0.082
Latent normal	−0.174	0.104	(−0.377, 0.030)	0.094

Table 5: Association between fibre intake and colerectal cancer estimated from EPIC-Norfolk. Log odds ratio is for six-gram per day increase in fibre intake, conditional on smoking, education, social class, physical activity, height, weight, age, alcohol intake, folate intake, intake of energy from fat and non-fat, aspirin use and the matching variables. Missing data are handled by restriction to complete cases or by MI.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
10% missing															
Complete cases	0.432	0.275	0.263	76.1	95	0.338	0.097	0.101	9.47	96	0.328	0.102	0.101	10.5	95
Match var: FCS	0.430	0.222	0.221	49.6	95	0.336	0.082	0.087	6.72	96	0.337	0.084	0.084	7.12	94
Normal	0.407	0.212	0.219	45.1	96	0.337	0.082	0.087	6.75	96	0.339	0.084	0.083	7.11	94
Latent norm	0.434	0.221	0.221	49.3	96	0.331	0.080	0.087	6.47	96	0.339	0.084	0.084	7.10	95
Match set: FCS	0.430	0.226	0.221	51.1	95	0.334	0.083	0.088	6.88	96	0.338	0.085	0.084	7.19	94
Normal	0.422	0.222	0.226	49.4	96	0.340	0.085	0.089	7.24	96	0.335	0.085	0.084	7.28	94
Latent norm	0.439	0.225	0.223	50.9	96	0.320	0.080	0.087	6.58	96	0.341	0.085	0.084	7.25	95
25% missing															
Complete cases	0.456	0.400	0.398	162	96	0.339	0.143	0.143	20.5	96	0.320	0.148	0.144	22.1	94
Match var: FCS	0.430	0.243	0.247	59.2	96	0.336	0.090	0.096	8.03	97	0.337	0.087	0.086	7.56	95
Normal	0.375	0.216	0.240	48.5	97	0.339	0.089	0.095	7.93	97	0.342	0.086	0.086	7.54	95
Latent norm	0.442	0.240	0.245	58.4	95	0.323	0.085	0.095	7.27	97	0.343	0.086	0.086	7.45	95
Match set: FCS	0.429	0.254	0.251	64.5	96	0.332	0.094	0.097	8.81	96	0.341	0.088	0.087	7.88	94
Normal	0.408	0.241	0.259	58.2	96	0.347	0.098	0.101	9.81	96	0.333	0.090	0.088	8.12	95
Latent norm	0.451	0.255	0.254	66.2	95	0.298	0.085	0.095	8.48	95	0.347	0.089	0.088	8.07	95

Web Table 1: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MAR-A missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
10% missing															
Complete cases	0.428	0.269	0.262	72.7	95	0.338	0.096	0.099	9.26	97	0.269	0.101	0.100	14.4	89
Match var: FCS	0.430	0.221	0.220	48.9	95	0.335	0.082	0.087	6.74	97	0.337	0.084	0.084	7.14	94
Normal	0.395	0.212	0.219	45.3	96	0.337	0.082	0.087	6.77	97	0.340	0.084	0.083	7.11	95
Latent norm	0.434	0.219	0.220	48.4	95	0.331	0.081	0.087	6.5	97	0.340	0.084	0.084	7.11	94
Match set: FCS	0.430	0.225	0.221	50.7	95	0.332	0.083	0.087	6.91	96	0.339	0.085	0.084	7.23	95
Normal	0.414	0.223	0.226	49.6	96	0.338	0.085	0.088	7.25	96	0.337	0.085	0.084	7.31	94
Latent norm	0.437	0.225	0.222	51	95	0.319	0.080	0.087	6.6	96	0.342	0.085	0.084	7.29	95
25% missing															
Complete cases	0.439	0.395	0.390	156	96	0.339	0.136	0.138	18.5	96	0.197	0.143	0.138	39.1	82
Match var: FCS	0.431	0.242	0.246	58.8	95	0.336	0.089	0.095	7.95	96	0.337	0.087	0.086	7.6	95
Normal	0.351	0.218	0.240	52	96	0.340	0.089	0.095	7.91	97	0.343	0.087	0.086	7.6	95
Latent norm	0.440	0.239	0.245	57.9	95	0.323	0.085	0.094	7.27	97	0.344	0.086	0.086	7.54	95
Match set: FCS	0.431	0.253	0.252	64.2	95	0.330	0.093	0.097	8.73	96	0.342	0.089	0.087	7.93	95
Normal	0.391	0.243	0.259	59.6	96	0.344	0.098	0.100	9.7	96	0.336	0.090	0.088	8.16	94
Latent norm	0.449	0.253	0.253	65	95	0.296	0.084	0.095	8.52	95	0.348	0.089	0.087	8.08	95

Web Table 2: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MAR-B missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.418	0.150	0.144	22.4	94	0.334	0.058	0.061	3.37	97	0.337	0.062	0.061	3.89	94
10% missing															
Complete cases	0.415	0.182	0.175	33.1	94	0.333	0.068	0.071	4.66	96	0.329	0.074	0.071	5.48	94
Match var: FCS	0.417	0.160	0.155	25.6	94	0.334	0.062	0.065	3.79	96	0.337	0.064	0.062	4.07	94
Normal	0.398	0.155	0.155	24.4	95	0.336	0.061	0.065	3.78	97	0.339	0.064	0.062	4.09	94
Latent norm	0.422	0.159	0.155	25.3	94	0.330	0.061	0.065	3.68	97	0.339	0.063	0.062	4.07	94
Match set: FCS	0.415	0.161	0.156	26.0	94	0.332	0.061	0.065	3.75	97	0.338	0.064	0.062	4.09	94
Normal	0.404	0.157	0.156	24.9	95	0.336	0.062	0.065	3.86	97	0.339	0.064	0.062	4.09	94
Latent norm	0.421	0.161	0.156	25.9	94	0.319	0.060	0.065	3.83	96	0.341	0.064	0.062	4.15	94
25% missing															
Complete cases	0.407	0.259	0.241	67.2	94	0.334	0.089	0.093	7.87	97	0.320	0.096	0.094	9.43	94
Match var: FCS	0.416	0.180	0.175	32.4	94	0.335	0.069	0.072	4.77	96	0.337	0.066	0.064	4.35	95
Normal	0.372	0.166	0.170	29.6	94	0.338	0.069	0.071	4.75	96	0.342	0.066	0.064	4.37	94
Latent norm	0.427	0.177	0.174	31.5	94	0.323	0.066	0.071	4.43	97	0.343	0.065	0.064	4.33	94
Match set: FCS	0.411	0.184	0.176	33.9	94	0.330	0.069	0.073	4.79	96	0.340	0.066	0.064	4.42	95
Normal	0.388	0.173	0.176	30.8	94	0.340	0.071	0.073	5.04	96	0.339	0.066	0.064	4.44	94
Latent norm	0.426	0.181	0.176	33.0	94	0.299	0.065	0.071	5.45	94	0.347	0.067	0.065	4.64	94

Web Table 3: Results from 1000 simulated datasets of $N = 500$ cases and $M = 4$ controls per case with MAR-A missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.418	0.150	0.144	22.4	94	0.334	0.058	0.061	3.37	97	0.337	0.062	0.061	3.89	94
10% missing															
Complete cases	0.407	0.184	0.176	33.8	94	0.334	0.067	0.071	4.55	96	0.266	0.073	0.071	9.9	83
Match var: FCS	0.416	0.160	0.156	25.5	94	0.334	0.061	0.065	3.76	97	0.337	0.064	0.062	4.09	94
Normal	0.378	0.155	0.155	25.3	95	0.336	0.061	0.065	3.75	97	0.340	0.064	0.062	4.11	94
Latent norm	0.419	0.158	0.156	25.1	95	0.329	0.060	0.065	3.64	97	0.340	0.064	0.062	4.09	94
Match set: FCS	0.415	0.161	0.157	26	95	0.332	0.061	0.065	3.74	97	0.339	0.064	0.062	4.14	94
Normal	0.388	0.156	0.157	25.3	95	0.335	0.062	0.065	3.81	97	0.340	0.064	0.062	4.13	94
Latent norm	0.416	0.159	0.157	25.4	95	0.319	0.060	0.065	3.8	96	0.342	0.064	0.062	4.21	94
25% missing															
Complete cases	0.397	0.262	0.243	69.1	94	0.334	0.087	0.092	7.65	97	0.190	0.096	0.092	29.8	64
Match var: FCS	0.415	0.182	0.177	33	95	0.335	0.069	0.072	4.77	97	0.337	0.066	0.064	4.36	94
Normal	0.335	0.168	0.172	34.7	93	0.340	0.069	0.072	4.78	96	0.344	0.066	0.064	4.42	94
Latent norm	0.421	0.178	0.176	31.7	95	0.323	0.065	0.072	4.36	97	0.343	0.065	0.064	4.33	95
Match set: FCS	0.411	0.188	0.180	35.5	94	0.330	0.070	0.074	4.89	96	0.340	0.067	0.065	4.53	94
Normal	0.355	0.173	0.177	33.7	94	0.339	0.071	0.073	5.02	96	0.342	0.066	0.065	4.47	95
Latent norm	0.417	0.181	0.179	32.9	96	0.298	0.066	0.072	5.55	95	0.348	0.067	0.065	4.73	94

Web Table 4: Results from 1000 simulated datasets of $N = 500$ cases and $M = 4$ controls per case with MAR-B missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.454	0.489	0.488	240	97	0.351	0.194	0.191	38	95	0.353	0.196	0.191	38.8	96
10% missing															
Complete cases	0.460	0.628	0.625	396	97	0.359	0.258	0.242	67.4	96	0.362	0.259	0.242	67.6	95
Match var: FCS	0.463	0.516	0.523	268	97	0.354	0.207	0.204	43.1	96	0.353	0.200	0.195	40.3	96
Normal	0.444	0.495	0.521	246	98	0.354	0.207	0.204	43.1	96	0.354	0.200	0.195	40.3	96
Latent norm	0.470	0.518	0.523	271	98	0.348	0.203	0.204	41.3	96	0.355	0.199	0.195	40.1	96
Match set: FCS	0.459	0.519	0.527	271	97	0.353	0.212	0.206	45.5	96	0.355	0.202	0.195	41.1	96
Normal	0.454	0.509	0.538	260	98	0.363	0.217	0.210	48.1	95	0.349	0.203	0.197	41.5	96
Latent norm	0.483	0.531	0.533	286	97	0.341	0.202	0.206	40.9	97	0.354	0.202	0.197	41.1	96
25% missing															
Complete cases	0.347	0.932	0.965	873	98	0.416	0.477	0.393	235	96	0.404	0.462	0.392	218	97
Match var: FCS	0.462	0.565	0.587	322	98	0.357	0.227	0.228	52.3	97	0.355	0.205	0.202	42.7	96
Normal	0.410	0.502	0.569	252	99	0.358	0.225	0.227	51	96	0.358	0.204	0.202	42.1	96
Latent norm	0.484	0.562	0.586	321	98	0.340	0.211	0.225	44.7	97	0.362	0.202	0.202	41.5	96
Match set: FCS	0.433	0.569	0.606	324	98	0.356	0.240	0.234	58.3	96	0.362	0.211	0.206	45.3	96
Normal	0.433	0.548	0.622	301	98	0.388	0.259	0.249	70	96	0.341	0.216	0.211	46.9	96
Latent norm	0.516	0.611	0.618	383	98	0.321	0.212	0.230	45.2	97	0.358	0.209	0.208	44.3	96

Web Table 5: Results from 1000 simulated datasets of $N = 100$ cases and $M = 1$ control per case with MCAR missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively. The complete-case analysis failed in 4 simulated datasets when 10% missing data, and in 67 datasets when 25% missing data. MI methods failed in no datasets when 10% missing data, and in at most 5 when 25% missing data. When calculating LOR, SE, estSE, MSE and cv for a method, datasets for which that method failed were excluded.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.454	0.489	0.488	240	97	0.351	0.194	0.191	38	95	0.353	0.196	0.191	38.8	96
10% missing															
Complete cases	0.458	0.655	0.647	431	97	0.360	0.250	0.240	63.5	96	0.351	0.256	0.241	65.9	95
Match var: FCS	0.461	0.523	0.530	275	97	0.357	0.205	0.204	42.6	96	0.351	0.199	0.195	39.9	96
Normal	0.435	0.495	0.526	246	98	0.358	0.205	0.204	42.6	96	0.353	0.199	0.195	39.9	96
Latent norm	0.467	0.521	0.530	274	97	0.351	0.201	0.204	40.5	96	0.354	0.198	0.195	39.8	96
Match set: FCS	0.464	0.531	0.534	284	97	0.354	0.207	0.205	43.4	95	0.355	0.200	0.195	40.5	96
Normal	0.459	0.521	0.548	274	97	0.365	0.215	0.210	47.2	96	0.348	0.202	0.197	41.1	96
Latent norm	0.481	0.540	0.541	296	96	0.341	0.200	0.206	39.9	96	0.354	0.200	0.197	40.6	96
25% missing															
Complete cases	0.317	0.973	0.997	957	98	0.408	0.422	0.366	184	96	0.369	0.422	0.367	179	97
Match var: FCS	0.468	0.603	0.605	366	97	0.359	0.224	0.228	50.8	96	0.353	0.205	0.203	42.5	96
Normal	0.409	0.534	0.584	285	98	0.360	0.222	0.226	49.8	97	0.358	0.204	0.202	42.1	96
Latent norm	0.487	0.595	0.600	359	97	0.342	0.207	0.225	43	97	0.360	0.202	0.203	41.4	97
Match set: FCS	0.458	0.603	0.630	366	98	0.357	0.236	0.235	56.3	96	0.361	0.211	0.206	45.2	96
Normal	0.454	0.572	0.647	329	99	0.387	0.254	0.247	67.1	95	0.342	0.215	0.210	46.2	96
Latent norm	0.517	0.649	0.643	431	98	0.321	0.209	0.230	43.7	97	0.359	0.208	0.208	44	96

Web Table 6: Results from 1000 simulated datasets of $N = 100$ cases and $M = 1$ control per case with MAR-A missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively. The complete-case analysis failed in 5 simulated datasets when 10% missing data, and in 82 datasets when 25% missing data. MI methods failed in no datasets when 10% missing data, and in at most 6 when 25% missing data. When calculating LOR, SE, estSE, MSE and cv for a method, datasets for which that method failed were excluded.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.454	0.489	0.488	240	97	0.351	0.194	0.191	38	95	0.353	0.196	0.191	38.8	96
10% missing															
Complete cases	0.444	0.655	0.641	430	97	0.360	0.249	0.236	63	96	0.291	0.257	0.236	67.7	95
Match var: FCS	0.458	0.521	0.528	273	97	0.355	0.206	0.203	43.1	96	0.352	0.200	0.195	40.3	96
Normal	0.422	0.499	0.524	249	98	0.356	0.205	0.203	42.5	96	0.355	0.199	0.194	40.3	96
Latent norm	0.464	0.520	0.528	273	97	0.350	0.200	0.203	40.5	96	0.354	0.199	0.195	39.9	97
Match set: FCS	0.465	0.528	0.533	281	97	0.351	0.208	0.205	43.7	95	0.356	0.201	0.195	40.9	96
Normal	0.453	0.518	0.547	269	97	0.360	0.215	0.209	47.1	95	0.351	0.202	0.197	41.3	96
Latent norm	0.475	0.536	0.540	291	97	0.338	0.199	0.205	39.6	96	0.355	0.201	0.196	40.9	96
25% missing															
Complete cases	0.326	0.958	0.978	927	98	0.403	0.407	0.351	170	96	0.239	0.406	0.350	174	93
Match var: FCS	0.465	0.589	0.604	349	97	0.360	0.227	0.227	52	96	0.350	0.208	0.203	43.4	96
Normal	0.379	0.529	0.583	282	98	0.361	0.223	0.225	50.7	96	0.357	0.205	0.202	42.7	96
Latent norm	0.479	0.591	0.599	353	98	0.342	0.210	0.224	44.4	97	0.359	0.204	0.203	42.2	96
Match set: FCS	0.481	0.617	0.633	385	98	0.354	0.242	0.235	59.1	96	0.358	0.214	0.207	46.3	96
Normal	0.452	0.579	0.647	336	98	0.380	0.256	0.245	67.9	95	0.345	0.217	0.209	47.3	96
Latent norm	0.513	0.651	0.643	433	98	0.320	0.211	0.228	44.7	97	0.358	0.211	0.208	45	96

Web Table 7: Results from 1000 simulated datasets of $N = 100$ cases and $M = 1$ control per case with MAR-B missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively. The complete-case analysis failed in 6 simulated datasets when 10% missing data, and in 88 datasets when 25% missing data. MI methods failed in at most one dataset when 10% missing data, and in at most 7 when 25% missing data. When calculating LOR, SE, estSE, MSE and cv for a method, datasets for which that method failed were excluded.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.402	0.334	0.330	112	95	0.336	0.143	0.138	20.4	94	0.339	0.135	0.138	18.4	96
10% missing															
Complete cases	0.408	0.387	0.388	150	96	0.340	0.165	0.160	27.3	95	0.341	0.161	0.160	26	96
Match var: FCS	0.400	0.349	0.350	122	95	0.336	0.147	0.146	21.7	96	0.339	0.137	0.140	18.7	96
Normal	0.391	0.340	0.349	116	97	0.338	0.147	0.146	21.7	96	0.340	0.136	0.140	18.6	96
Latent norm	0.406	0.349	0.350	122	96	0.332	0.145	0.146	20.9	96	0.341	0.136	0.140	18.5	96
Match set: FCS	0.390	0.347	0.349	121	96	0.332	0.145	0.146	21.1	96	0.342	0.136	0.140	18.5	96
Normal	0.395	0.345	0.353	119	96	0.340	0.148	0.147	22	96	0.339	0.137	0.141	18.8	96
Latent norm	0.408	0.352	0.352	124	95	0.325	0.144	0.146	20.9	96	0.341	0.138	0.141	19	96
25% missing															
Complete cases	0.437	0.564	0.533	318	95	0.353	0.219	0.216	48.3	96	0.348	0.219	0.216	48.1	96
Match var: FCS	0.405	0.388	0.388	151	96	0.337	0.161	0.162	26	96	0.339	0.139	0.145	19.3	96
Normal	0.380	0.360	0.379	131	97	0.341	0.160	0.161	25.7	96	0.342	0.138	0.144	19.2	96
Latent norm	0.421	0.388	0.387	150	96	0.324	0.153	0.160	23.5	97	0.345	0.137	0.145	19	96
Match set: FCS	0.375	0.375	0.386	142	96	0.325	0.156	0.160	24.4	97	0.349	0.138	0.144	19.2	96
Normal	0.392	0.372	0.393	139	96	0.347	0.168	0.166	28.4	95	0.337	0.142	0.146	20.1	95
Latent norm	0.429	0.400	0.395	161	96	0.308	0.154	0.162	24.3	96	0.344	0.142	0.147	20.2	96

Web Table 8: Results from 1000 simulated datasets of $N = 100$ cases and $M = 4$ controls per case with MCAR missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.402	0.334	0.330	112	95	0.336	0.143	0.138	20.4	94	0.339	0.135	0.138	18.4	96
10% missing															
Complete cases	0.406	0.407	0.402	166	95	0.341	0.168	0.162	28.1	95	0.329	0.159	0.162	25.5	96
Match var: FCS	0.402	0.357	0.356	127	96	0.338	0.149	0.147	22.3	96	0.339	0.136	0.141	18.6	96
Normal	0.385	0.347	0.354	122	96	0.339	0.149	0.147	22.3	96	0.340	0.136	0.140	18.6	95
Latent norm	0.406	0.357	0.356	128	96	0.333	0.147	0.147	21.5	96	0.341	0.136	0.141	18.5	96
Match set: FCS	0.396	0.351	0.355	124	96	0.332	0.148	0.147	21.8	96	0.342	0.136	0.141	18.5	96
Normal	0.395	0.349	0.359	122	96	0.340	0.150	0.148	22.6	96	0.340	0.136	0.141	18.6	96
Latent norm	0.407	0.358	0.358	129	96	0.325	0.146	0.147	21.4	96	0.342	0.137	0.141	18.8	96
25% missing															
Complete cases	0.436	0.616	0.577	380	95	0.353	0.218	0.215	47.8	96	0.331	0.220	0.217	48.5	96
Match var: FCS	0.399	0.405	0.405	165	95	0.338	0.161	0.163	26	96	0.338	0.140	0.146	19.7	96
Normal	0.359	0.369	0.393	140	96	0.341	0.160	0.163	25.6	96	0.343	0.139	0.145	19.4	96
Latent norm	0.412	0.401	0.401	161	95	0.325	0.153	0.161	23.4	97	0.345	0.138	0.146	19.3	96
Match set: FCS	0.378	0.390	0.403	153	97	0.321	0.155	0.163	24.2	97	0.349	0.139	0.146	19.6	96
Normal	0.385	0.384	0.409	149	97	0.345	0.167	0.167	28	96	0.339	0.142	0.147	20.3	96
Latent norm	0.417	0.417	0.411	174	95	0.306	0.153	0.163	24.2	96	0.346	0.142	0.148	20.4	96

Web Table 9: Results from 1000 simulated datasets of $N = 100$ cases and $M = 4$ controls per case with MAR-A missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
Complete data	0.402	0.334	0.330	112	95	0.336	0.143	0.138	20.4	94	0.339	0.135	0.138	18.4	96
10% missing															
Complete cases	0.395	0.420	0.406	177	95	0.341	0.167	0.161	28	94	0.265	0.158	0.162	29.6	93
Match var: FCS	0.398	0.364	0.358	133	95	0.338	0.150	0.147	22.6	96	0.338	0.136	0.141	18.6	95
Normal	0.363	0.352	0.356	127	96	0.340	0.149	0.147	22.4	96	0.341	0.136	0.141	18.5	95
Latent norm	0.401	0.362	0.357	132	96	0.333	0.148	0.147	21.9	96	0.341	0.136	0.141	18.5	96
Match set: FCS	0.396	0.360	0.357	130	96	0.331	0.149	0.147	22.1	96	0.342	0.136	0.141	18.6	96
Normal	0.378	0.354	0.361	127	96	0.338	0.151	0.148	22.9	96	0.341	0.136	0.141	18.7	96
Latent norm	0.399	0.368	0.360	135	95	0.325	0.147	0.148	21.6	96	0.342	0.137	0.142	18.8	96
25% missing															
Complete cases	0.403	0.624	0.583	390	95	0.354	0.216	0.212	47.3	96	0.193	0.216	0.213	66.3	89
Match var: FCS	0.391	0.411	0.411	169	96	0.340	0.162	0.164	26.3	96	0.336	0.142	0.146	20.1	96
Normal	0.317	0.373	0.398	149	96	0.344	0.162	0.163	26.3	96	0.343	0.141	0.145	20	95
Latent norm	0.398	0.406	0.408	165	96	0.327	0.154	0.163	23.9	97	0.343	0.140	0.146	19.7	96
Match set: FCS	0.386	0.396	0.411	158	96	0.322	0.157	0.166	24.6	96	0.346	0.141	0.147	20.1	96
Normal	0.355	0.384	0.413	151	97	0.343	0.167	0.167	28.1	95	0.342	0.144	0.147	20.7	96
Latent norm	0.399	0.421	0.418	177	96	0.307	0.155	0.164	24.8	96	0.346	0.144	0.149	20.9	95

Web Table 10: Results from 1000 simulated datasets of $N = 100$ cases and $M = 4$ controls per case with MAR-B missingness mechanism. 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
10% missing															
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
Complete cases	0.431	0.264	0.256	70	96	0.337	0.100	0.102	10.1	96	0.340	0.103	0.102	10.6	96
Match var: FCS	0.432	0.223	0.219	50.1	95	0.332	0.083	0.087	6.9	96	0.338	0.085	0.084	7.2	94
Normal	0.412	0.214	0.218	45.6	96	0.333	0.083	0.087	6.82	96	0.340	0.085	0.084	7.2	95
Latent norm	0.437	0.222	0.218	49.6	96	0.327	0.081	0.087	6.64	96	0.340	0.084	0.084	7.17	95
Match set: FCS	0.429	0.226	0.219	51.1	95	0.334	0.084	0.088	7.09	96	0.338	0.085	0.084	7.23	94
Normal	0.422	0.222	0.223	49.2	96	0.340	0.086	0.089	7.49	96	0.335	0.086	0.084	7.35	95
Latent norm	0.445	0.225	0.220	51.3	95	0.311	0.081	0.087	7.01	95	0.342	0.086	0.084	7.44	95
25% missing															
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
Complete cases	0.449	0.379	0.377	145	96	0.341	0.144	0.149	20.8	97	0.342	0.150	0.149	22.4	96
Match var: FCS	0.439	0.239	0.241	57.8	96	0.328	0.089	0.095	7.98	96	0.339	0.087	0.086	7.59	95
Normal	0.390	0.215	0.235	46.8	97	0.330	0.089	0.095	7.93	97	0.342	0.087	0.086	7.58	95
Latent norm	0.451	0.238	0.240	58	96	0.316	0.084	0.095	7.45	96	0.345	0.086	0.086	7.5	95
Match set: FCS	0.430	0.247	0.243	61.3	95	0.335	0.094	0.098	8.87	96	0.339	0.088	0.087	7.71	95
Normal	0.413	0.239	0.251	57	97	0.351	0.099	0.102	10.1	95	0.329	0.090	0.088	8.12	94
Latent norm	0.476	0.249	0.246	65.7	95	0.279	0.084	0.095	9.98	93	0.348	0.090	0.089	8.3	95

Web Table 11: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MCAR missingness mechanism in Sensitivity Analysis 1 (where there is an interaction between S_{cat} and S_{con}). 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
10% missing															
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
Complete cases	0.432	0.275	0.264	76.1	95	0.338	0.097	0.101	9.42	96	0.328	0.102	0.101	10.5	95
Match var: FCS	0.434	0.222	0.221	49.5	95	0.332	0.081	0.087	6.6	96	0.338	0.084	0.084	7.13	94
Normal	0.408	0.212	0.220	45.1	96	0.334	0.081	0.087	6.64	96	0.340	0.084	0.084	7.12	95
Latent norm	0.438	0.220	0.221	48.9	96	0.328	0.080	0.087	6.45	96	0.340	0.084	0.084	7.1	95
Match set: FCS	0.431	0.227	0.222	51.7	95	0.334	0.083	0.088	6.85	96	0.339	0.084	0.084	7.15	95
Normal	0.424	0.222	0.226	49.5	95	0.339	0.084	0.089	7.17	96	0.336	0.085	0.084	7.24	94
Latent norm	0.448	0.225	0.223	51.8	95	0.310	0.079	0.087	6.82	96	0.343	0.085	0.084	7.34	95
25% missing															
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
Complete cases	0.457	0.408	0.397	168	96	0.339	0.140	0.141	19.7	96	0.321	0.147	0.143	21.6	94
Match var: FCS	0.437	0.244	0.248	59.8	96	0.328	0.089	0.095	8.02	96	0.339	0.088	0.086	7.7	95
Normal	0.377	0.217	0.240	48.7	97	0.331	0.089	0.095	7.9	97	0.344	0.087	0.086	7.71	95
Latent norm	0.448	0.241	0.246	58.9	96	0.316	0.085	0.094	7.54	96	0.345	0.087	0.086	7.64	95
Match set: FCS	0.430	0.256	0.253	65.5	96	0.330	0.094	0.098	8.92	96	0.342	0.089	0.087	7.93	95
Normal	0.410	0.242	0.260	58.7	97	0.347	0.100	0.101	10.1	96	0.333	0.091	0.088	8.24	94
Latent norm	0.471	0.255	0.254	68.1	94	0.277	0.085	0.095	10.3	93	0.351	0.090	0.088	8.47	94

Web Table 12: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MAR-A missingness mechanism in Sensitivity Analysis 1 (where there is an interaction between S_{cat} and S_{con}). 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv	LOR	SE	estSE	MSE	cv
10% missing															
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
Complete cases	0.427	0.272	0.262	73.9	94	0.336	0.096	0.099	9.16	97	0.272	0.101	0.099	13.9	88
Match var: FCS	0.435	0.220	0.221	48.8	95	0.331	0.082	0.087	6.68	97	0.338	0.085	0.084	7.2	95
Normal	0.394	0.211	0.219	45.1	96	0.333	0.082	0.087	6.7	97	0.342	0.084	0.084	7.17	95
Latent norm	0.439	0.218	0.220	48.2	96	0.326	0.080	0.087	6.51	97	0.341	0.084	0.084	7.18	95
Match set: FCS	0.434	0.225	0.222	50.8	94	0.330	0.083	0.087	6.94	96	0.340	0.085	0.084	7.27	94
Normal	0.414	0.222	0.226	49.2	96	0.336	0.085	0.088	7.26	96	0.338	0.085	0.084	7.31	94
Latent norm	0.448	0.222	0.223	50.2	95	0.307	0.080	0.087	7.07	95	0.345	0.086	0.084	7.49	94
25% missing															
Complete data	0.426	0.213	0.206	45.3	94	0.336	0.078	0.082	6.15	96	0.337	0.083	0.082	6.88	95
Complete cases	0.428	0.318	0.302	101	95	0.339	0.109	0.111	11.9	96	0.244	0.114	0.111	21.1	86
Match var: FCS	0.438	0.228	0.230	52.4	96	0.329	0.085	0.089	7.16	96	0.339	0.086	0.085	7.39	95
Normal	0.375	0.212	0.227	46.7	96	0.333	0.084	0.089	7.06	97	0.343	0.085	0.084	7.37	95
Latent norm	0.442	0.226	0.230	51.6	97	0.322	0.082	0.089	6.82	97	0.343	0.085	0.085	7.33	95
Match set: FCS	0.435	0.236	0.233	56.1	95	0.329	0.087	0.091	7.65	96	0.341	0.087	0.085	7.55	95
Normal	0.407	0.229	0.239	52.4	96	0.339	0.090	0.092	8.1	96	0.338	0.088	0.085	7.68	95
Latent norm	0.458	0.233	0.234	56.1	95	0.293	0.081	0.089	8.2	94	0.349	0.087	0.086	7.85	95

Web Table 13: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MAR-B missingness mechanism in Sensitivity Analysis 1 (where there is an interaction between S_{cat} and S_{con}). 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	eSE	estSE	MSE	cv	LOR	eSE	estSE	MSE	cv	LOR	eSE	estSE	MSE	cv
10% missing															
Complete data	0.414	0.198	0.195	39.2	94	0.336	0.073	0.075	5.3	96	0.338	0.078	0.075	6.07	94
Complete cases	0.421	0.245	0.242	60	95	0.340	0.092	0.093	8.55	94	0.342	0.098	0.093	9.77	94
Match var: FCS	0.428	0.204	0.207	41.7	96	0.326	0.071	0.078	5.03	96	0.345	0.078	0.076	6.26	94
Normal	0.408	0.195	0.206	38.1	96	0.327	0.071	0.078	5.07	97	0.345	0.078	0.076	6.29	95
Latent norm	0.420	0.205	0.207	42	96	0.337	0.073	0.079	5.41	97	0.340	0.079	0.076	6.22	95
Match set: FCS	0.428	0.206	0.208	42.7	96	0.325	0.072	0.078	5.21	96	0.345	0.078	0.076	6.28	95
Normal	0.417	0.202	0.211	41	96	0.330	0.072	0.079	5.18	97	0.343	0.079	0.076	6.3	95
Latent norm	0.413	0.211	0.210	44.5	95	0.337	0.075	0.080	5.65	97	0.338	0.079	0.076	6.32	95
25% missing															
Complete data	0.414	0.198	0.195	39.2	94	0.336	0.073	0.075	5.3	96	0.338	0.078	0.075	6.07	94
Complete cases	0.423	0.353	0.356	124	96	0.350	0.136	0.136	18.8	96	0.350	0.141	0.137	20.3	95
Match var: FCS	0.443	0.222	0.228	49.9	96	0.314	0.072	0.084	5.6	96	0.353	0.080	0.078	6.77	95
Normal	0.394	0.202	0.223	41.3	97	0.316	0.073	0.084	5.59	97	0.356	0.080	0.078	6.86	95
Latent norm	0.419	0.231	0.229	53.4	95	0.338	0.079	0.086	6.25	97	0.343	0.080	0.078	6.57	95
Match set: FCS	0.441	0.231	0.232	53.9	96	0.312	0.075	0.084	6.12	96	0.355	0.080	0.078	6.87	95
Normal	0.414	0.221	0.238	48.6	97	0.325	0.077	0.087	6.08	97	0.348	0.081	0.079	6.83	95
Latent norm	0.410	0.252	0.241	63.3	95	0.338	0.084	0.090	7	97	0.338	0.083	0.079	6.88	95

Web Table 14: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MCAR missingness mechanism in Sensitivity Analysis 2 (where X^{conA} and X^{conB} are log normally distributed). 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	eSE	estSE	MSE	cv	LOR	eSE	estSE	MSE	cv	LOR	eSE	estSE	MSE	cv
10% missing															
Complete data	0.414	0.198	0.195	39.2	94	0.336	0.073	0.075	5.3	96	0.338	0.078	0.075	6.07	94
Complete cases	0.422	0.245	0.246	59.9	95	0.341	0.092	0.092	8.55	96	0.328	0.103	0.098	10.6	94
Match var: FCS	0.427	0.204	0.208	41.8	96	0.330	0.073	0.079	5.41	96	0.336	0.080	0.077	6.45	94
Normal	0.404	0.195	0.208	38.3	97	0.331	0.073	0.079	5.36	96	0.338	0.080	0.077	6.45	95
Latent norm	0.419	0.204	0.208	41.5	96	0.339	0.076	0.079	5.81	96	0.332	0.081	0.077	6.49	94
Match set: FCS	0.409	0.207	0.209	42.7	96	0.323	0.074	0.079	5.65	96	0.341	0.081	0.078	6.67	94
Normal	0.417	0.203	0.213	41.1	96	0.332	0.075	0.080	5.6	97	0.335	0.081	0.078	6.59	94
Latent norm	0.413	0.209	0.211	43.9	95	0.337	0.077	0.080	5.92	96	0.331	0.082	0.078	6.7	94
25% missing															
Complete data	0.414	0.198	0.195	39.2	94	0.336	0.073	0.075	5.3	96	0.338	0.078	0.075	6.07	94
Complete cases	0.427	0.366	0.366	134	96	0.351	0.132	0.133	17.7	95	0.316	0.149	0.150	22.5	94
Match var: FCS	0.438	0.227	0.232	52	95	0.322	0.076	0.086	5.89	97	0.335	0.084	0.081	7.06	95
Normal	0.384	0.206	0.227	43.6	97	0.324	0.076	0.086	5.81	98	0.339	0.084	0.081	7.03	95
Latent norm	0.421	0.234	0.233	54.8	95	0.341	0.082	0.087	6.71	97	0.326	0.085	0.081	7.3	94
Match set: FCS	0.398	0.238	0.236	56.8	95	0.307	0.078	0.086	6.83	95	0.344	0.087	0.082	7.63	95
Normal	0.415	0.226	0.243	51.1	97	0.327	0.080	0.089	6.45	97	0.329	0.086	0.082	7.46	95
Latent norm	0.415	0.253	0.244	63.8	94	0.335	0.085	0.090	7.2	97	0.320	0.088	0.083	7.99	93

Web Table 15: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MAR-A missingness mechanism in Sensitivity Analysis 2 (where X^{conA} and X^{conB} are log normally distributed). 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}					X^{conA}					X^{conB}				
	LOR	eSE	estSE	MSE	cv	LOR	eSE	estSE	MSE	cv	LOR	eSE	estSE	MSE	cv
10% missing															
Complete data	0.414	0.198	0.195	39.2	94	0.336	0.073	0.075	5.3	96	0.338	0.078	0.075	6.07	94
Complete cases	0.431	0.238	0.237	56.8	95	0.350	0.091	0.090	8.57	95	0.216	0.106	0.098	25	73
Match var: FCS	0.422	0.202	0.206	40.9	96	0.344	0.077	0.079	5.99	96	0.321	0.083	0.079	7.09	93
Normal	0.393	0.196	0.205	38.9	96	0.345	0.076	0.079	5.97	96	0.324	0.083	0.079	7.01	94
Latent norm	0.416	0.202	0.206	40.8	96	0.351	0.079	0.080	6.48	95	0.318	0.084	0.079	7.22	93
Match set: FCS	0.414	0.205	0.207	41.9	96	0.333	0.076	0.079	5.7	96	0.325	0.085	0.080	7.36	93
Normal	0.410	0.202	0.210	41	97	0.343	0.078	0.080	6.1	95	0.320	0.085	0.080	7.34	93
Latent norm	0.410	0.206	0.209	42.6	96	0.347	0.079	0.080	6.46	95	0.315	0.085	0.080	7.59	92
25% missing															
Complete data	0.414	0.198	0.195	39.2	94	0.336	0.073	0.075	5.3	96	0.338	0.078	0.075	6.07	94
Complete cases	0.437	0.341	0.347	117	96	0.355	0.125	0.126	16.1	96	0.105	0.152	0.148	75.4	64
Match var: FCS	0.426	0.226	0.231	51	96	0.346	0.081	0.088	6.79	97	0.310	0.088	0.084	8.27	93
Normal	0.353	0.206	0.225	46.4	96	0.349	0.081	0.088	6.88	97	0.317	0.087	0.083	7.9	93
Latent norm	0.412	0.228	0.230	51.9	95	0.360	0.086	0.089	8.04	96	0.303	0.089	0.084	8.76	92
Match set: FCS	0.409	0.234	0.236	55	95	0.324	0.080	0.088	6.56	96	0.317	0.094	0.086	9.15	93
Normal	0.398	0.225	0.241	50.9	96	0.346	0.085	0.091	7.43	97	0.306	0.091	0.086	9	92
Latent norm	0.410	0.241	0.239	58	95	0.349	0.086	0.090	7.71	96	0.296	0.092	0.086	9.9	91

Web Table 16: Results from 1000 simulated datasets of $N = 500$ cases and $M = 1$ control per case with MAR-B missingness mechanism in Sensitivity Analysis 2 (where X^{conA} and X^{conB} are log normally distributed). 'LOR' is mean estimated log odds ratio, 'SE' is empirical standard error, 'estSE' is mean estimated standard error, 'MSE' is mean squared error $\times 1000$, and 'cv' is coverage of 95% confidence interval. True log odds ratios are 0.417, 0.333 and 0.333 for X^{cat}, X^{conA} and X^{conB}, respectively.

	X^{cat}		X^{conA}		X^{conB}	
Method	Est	SE	Est	SE	Est	SE
Complete cases	0.443	0.211	0.235	0.089	0.269	0.083
MI using matching variables						
FCS	0.306	0.159	0.305	0.069	0.314	0.059
Normal	0.298	0.151	0.318	0.067	0.310	0.059
Latent norm	0.324	0.158	0.300	0.072	0.316	0.060
MI using matched set						
FCS	0.289	0.155	0.310	0.074	0.313	0.061
Normal	0.290	0.157	0.312	0.072	0.312	0.060
Latent norm	0.321	0.165	0.299	0.076	0.307	0.062

Web Table 17: Results from the specimen simulated dataset of Appendix H. Dataset has $N = 500$ cases and $M = 4$ controls per case with 25% missing values in X^{cat} and X^{conA}, generated independently in each variable and completely at random. 'Est' is the estimated log odds ratio and 'SE' is its estimated standard error.