### Pooling multiple imputations when the sample happens to be the population. (Vink 2014)

### Toward a standardized evaluation of imputation methodology. (Hanne,Vink 2023)

#### 2.3 Missingness generation

Missingness should be induced according to several sets of missing data conditions. We encourage simulators to consider different missingness patterns and mechanisms.

Truly random missingness across all data entries (“missing completely at random” or MCAR under Rubin’s (1976) definition) may be considered as a necessary simulation condition for the evaluation of imputation procedures, ….

If an imputation method is not able to solve the problem (i.e., yield valid inference) under MCAR, the statistical properties of the procedure are not universally sound.

A straightforward technique for inducing univariate MAR missingness is described in Van Buuren (2018, §3.2.4), whereas generalizations to multivariate MAR missingness can be found in Schouten et al. (2018).

The effects of different types of MAR mechanisms are described in Schouten and Vink (2021).

#### 2.4 Performance evaluation

We therefore recommend evaluating the following points:

1. The methods should preferably be unbiased. Absolute or relative bias calculations.
2. The intervals around estimates should have a valid coverage of the population (i.e., true) value. Coverage of a 95% interval should in theory be ≥ 95%, where a coverage rate of 95% would be most efficient.
3. The width of the confidence or credible interval may convey statistical efficiency, which should be considered to compare imputation methods.
4. Resemblance to the true data values may be quantified using the root mean squared error (RMSE). RMSE not recommend except when the aim is prediction.

Even though the estimation on the analysis level may be justified, some methods can yield imputations that may seem completely invalid to applied researchers. For example, one could very accurately estimate average human height by filling in negative values and values that are unrealistically large. While the obtained inference could still be valid under such imputations, the plausibility of the imputed values given the observed data should be under scrutiny. …. Many techniques can yield valid inferences, but techniques that sample realistic or plausible values may be preferable in practice.

Every imputation workflow should therefore contain an evaluation of the obtained imputations. Even though inspecting each imputation may be labor-intensive due to the number of imputations generated in a simulation study, we highly recommend simulators consider the following aspects:

1. The absence of nonconvergence in the imputation-generating process is a minimum requirement for any imputation method.
2. The fit of the imputation model may be verified with the help of a posterior predictive check. A straightforward posterior predictive check for imputation methodology is the multiple overimputation of observed data values.
3. The distributional characteristics of the imputations should be inspected for anomalies. The distribution of the incomplete data may differ greatly from the observed data.
4. Finally, the plausibility of the imputed values may be evaluated. Plausible imputations are not necessary for obtaining valid inference, but may be a desired property.

We know the theoretical properties of complete case analysis, which makes the ad hoc technique useful as a lower limit for evaluating imputation performance. Complete case analysis may therefore serve as a benchmark method against which imputation performance should be evaluated. Realistically, the simulator would expect the imputation solution to perform not worse than the complete case analysis and, preferably, mimic the complete data results.

#### 3 | SUGGESTED COURSE OF ACTION

The simulation design could be presented textually, in a flowchart or as a block of pseudocode, whereas missingness mechanisms could be written as a function of the data or displayed graphically. Ideally, the evaluations should be supplemented by an online repository with all of the data and code required to reproduce the simulation results. To aid simulators in reporting and move toward standardization in evaluation, we provide a draft version for reporting guidelines in Appendix A.1 (also available from <https://www.gerkovink.com/evaluation/>).

### Generating missing values for simulation purposes: a multivariate amputation procedure. (Schouten et al. 2018)

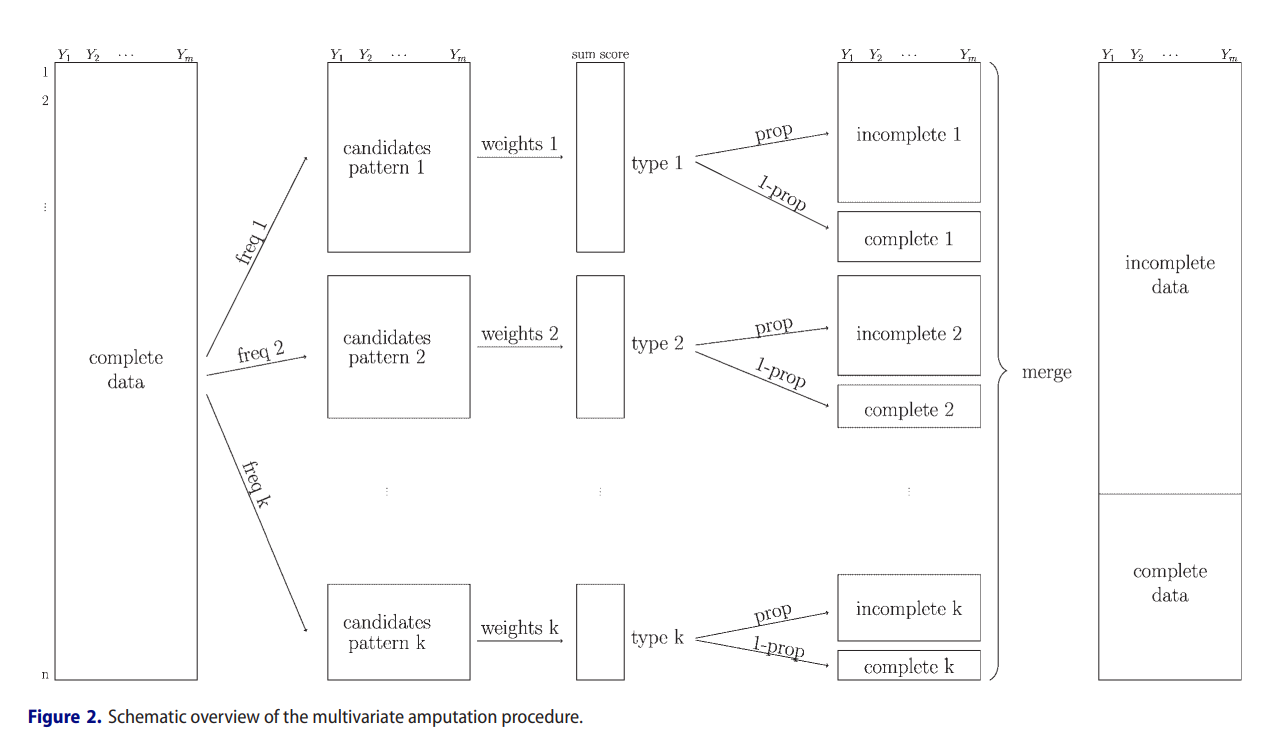
#### 1. Introduction

In current simulation studies, missing values are generated one variable at a time. With this univariate amputation approach, it can be difficult to appropriately control the characteristics of a missing data problem. ….

As a result, missing data methods may be evaluated under the wrong conditions and invalid conclusions about their inferential performance may be drawn.

Because our amputation procedure generates missing values in a multivariate way, issues with unreliable and inconsistent missing data generation can be overcome. Additionally, to make the multivariate amputation procedure available to a broad audience, we implemented the methodology as the function *ampute* in R-package mice.

#### 3.1. Multivariate amputation



The weighted sum score of case i is calculated as follows:

*wssi = w1 · y1i + w2 · y2i +···+ wm · ymi*

where {y1i, y2i, ... , ymi} is the set of variable values of case i and {w1,w2, ... ,wm} are the corresponding pre-specified weights.

In the last step of the multivariate amputation procedure, the assigned probabilities are used to generate missing values. Because we use a probabilistic model, some cases will remain complete, while other will receive missing values.

#### 3.2. Benefits of multivariate amputation

By offering the possibility to specify weight values, the multivariate amputation procedure also enables the fine-tuning of missing data mechanisms. Further, all aspects of the multivariate amputation methodology as discussed above can be regulated for different types of data (categorical data will be made numeric). Moreover, a smart use of the weight matrix enables the generation of both MAR and MNAR mechanisms within the same data set.

These and other features of ampute ensure the generation of complex – but realistic – missing data problems. For a detailed explanation about ampute’s arguments, we refer to [this tutorial](https://rianneschouten.github.io/mice_ampute/vignette/ampute.html).

### The Dance of the Mechanisms: How Observed Information Influences the Validity of Missingness Assumptions. (Schouten & Vink 2021)

#### Missingness Assumptions

In this context, Rubin (1976) distinguished three types of probability distributions: MCAR, MAR, and MNAR missingness.

#### Design of the Experiment

The design of the simulation basically consists of four steps: data set sampling with changing data correlations, generating missing values with varying missingness proportions and missingness mechanisms, each replicated 1000 times.

On those samples use MI to impute values and then use evaluation of statistical parameters.

missingness mechanisms ← {mcar; marright; marleft; marmid; martail mnarright; mnarleft;

mnarmid; mnartail}

correlations ← {0.1, 0.2, ... , 0.9}

missingness proportions ← {0.1, 0.5, 0.9}

replications ← 1000

#### Key Findings

##### Proportions

Thus, whether or not our missingness assumptions are valid is in essence not determined by the missingness proportion. However, an increased missingness proportion does have an effect on the size of possible bias, coverage, and variance measures.

##### Mean

* With low true data correlations, CCA estimates of EY are comparable between MCAR and MAR mechanisms.
* With increasing data correlations, MI uses the information in observed variable X to obtain unbiased estimates of EY. Even under MNAR missingness, MI uses the information in X to decrease the bias in estimates of EY

##### Variance

* Regardless of whether the underlying mechanism is MAR or MNAR, MID, and TAIL, missingness influences the variance of the observed data distribution.
* High data correlations enable MI to reduce the variance in estimates of EY , even for MCAR missingness.

##### Correlation

* The estimated correlation coefficient is not affected by MAR and MCAR missingness.
* With low and high true data correlations, CCA estimates of BX under MNAR approach the values of the estimates obtained under MCAR and MAR.
* Despite the size of data correlations, MI does not affect the estimates of BX.

#### Discussion

When the variables of a data set are barely correlated, MAR and MCAR missingness may become indistinguishable. In such cases, assuming MAR to allow for the use of MI would limit statistical power. MI would merely increase the variance without the need for decreasing bias.

Assuming MNAR missingness on highly correlated data may be unnecessary as the essence of the missing data may be sufficiently covered by the observed information.

….. As a result, an applied researcher’s hope for obtaining valid MAR-based inference on MNAR data can be justified when their variables are highly correlated. It is needless to say we can “force” such a situation by including the right variables.

When performing simulation studies, we strongly recommend researchers to perform CCA to investigate the effects of a generated missing data problem. When CCA returns biases close to zero, coverage rates close to 95 percent or very large ciws, it is wise to revisit the simulation conditions.