**1 Introduction**

Multilevel data are clustered data structures, for example with students clustered in classes or patients clustered in hospitals. Such hierarchical data structures may be accommodated using multilevel modeling—a statistical technique to capture associations in the data both at the unit-level (level-1) as well as the cluster-level (level-2; REF: Hox). Multilevel modeling is a step-by-step procedure (cf. Hox et al., 2017). First, a null-model is fitted to serve as a benchmark. Then, level-1 effects, level-2 effects and cross-level interactions can be added sequentially. The procedure of multilevel modeling can be complicated, especially in the presence of missing values.

If there are missing entries in multilevel data, even the null-model cannot be estimated. That’s why statistical software often defaults to ignoring any incomplete rows in the data. This method, known as ‘list-wise deletion’, allows the analyst to estimate statistical models on the subset of complete cases in the data. Ignoring incomplete cases assumes that the complete cases are representative of the entire sample. If there is a difference between complete and incomplete cases, complete-case analysis will yield biased results (REF: van Buuren). Moreover, with multilevel data structures, missingness can occur at the unit-level (i.e., incomplete cases), but also at the cluster-level (i.e., incomplete or entirely unobserved cluster-level variables). If the observations for a variable are systematically missing in a certain cluster, none of the units in this cluster can be included in the analysis of scientific interest after list-wise deletion. Complete-case analysis then is a wasteful way to accommodate missing data, and may even further bias the analysis results. Hence, the default method to handle missingness in multilevel data is not appropriate.

A statistically valid way to handle incomplete data is to impute (i.e., fill in) every missing data entry multiple times. With multiple imputation, several completed versions of the incomplete data are created, which can be analyzed as if the data were complete. The resulting statistics can then be pooled according to Rubin’s rules (REF: Rubin). Since the analysis results may vary across imputations, pooled estimates will reflect the uncertainty due to non-response. Multiple imputation has been established as a valid all-round method to deal with incomplete data.

This tutorial focuses on one ‘flavor’ of multiple imputation: multiple imputation by chained equations (MICE) as implemented in the R package {mice}. MICE is an algorithmic approach to fill in missing data entries on a variable-by-variable basis. That means that for every incomplete variable in the data, an imputation model should be chosen. This imputation model should be congenial (i.e., statistically matching) the analysis model one is intending to fit after treating the missingness. There is no encompassing imputation model (‘joint model’) required. The flexibility that this approach gives, may also be a disadvantage. If it is not clear at the imputation stage what the final analysis model will be at the multilevel modeling stage, it may be difficult to select appropriate imputation models. Each imputation model should therefore be compatible with the broadest multilevel model the analyst is intending to fit after imputation. In this tutorial, we will outline how to build imputation models that are in line with the multilevel structure in the data.

The aim of this tutorial is to provide empirical researchers who are faced with incomplete multilevel data guidance in imputing the missing values in their data. Missing data pose a wicked, but treatable problem. Combined with multilevel structures (another wicked but treatable problem), the solutions are a lot more difficult. This tutorial will aid empirical researchers in validly handling incomplete multilevel data by means of multiple imputation with the R package {mice}. Other valid methods (such as smcfcs) are outside the scope of this tutorial. No experience with missing data and imputation is required. We will assume basic familiarity with multilevel modeling. The notation will follow Hox et al. (2017). The R package {lme4} is used for the analysis of scientific interest. Analysis estimates after imputation are pooled with {miceadds}. Visualizations for data exploration and evaluation of the imputations will be performed using {ggplot2} and {ggmice}.

The requirements to follow along with this tutorial are an incomplete dataset with a multilevel structure, and the conviction that there are no reasons for the missingness that cannot be modeled from the data (ignorability assumption, see Box XYZ). Software-wise, this tutorial requires R and the packages mice, lme4, ggplot2, ggmice and miceadds. As a case study, we will use an adapted version of the popularity data, published by Hox et al. (2017). The original dataset does not contain any missing values. For the purpose of this tutorial, we have created an ‘amputed’ (i.e. incomplete) version of the data. The complete version will serve as comparative truth in this tutorial. Both the complete and incomplete data can be downloaded from XYZ. By the end of this tutorial, the reader will be able to build an imputation model for each incomplete variable, evaluate the imputation models, impute the data, evaluate the imputations, and analyze the data, and evaluate the effects of the imputations.