**Imputation of Incomplete Multilevel Data with R**

Oberman, H.I.1\*, Muñoz, J.2, Vink G.1, Debray T.P.A.2, de Jong, V.M.T.2,3

1 Methodology & Statistics, Utrecht University, The Netherlands

2 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands

3 Data Analytics and Methods Task Force, European Medicine Agency, The Netherlands

\* Presenting author (PhD candidate, supervisors: Gerko Vink en Stef van Buuren)

**Suggested talk duration (15-60 minutes)**

Depending on the conference organization, our proposal would be suitable either for a short paper presentation (15 minutes), or a longer interactive tutorial session (45-60 minutes).

**Summary (max. 500 words)**

Incomplete multilevel data requires careful consideration of the missing data problem and analysis strategy. In this tutorial, we focus on a popular strategy for accommodating missingness in multilevel data: replacing the missing data with plausible values, i.e., imputation.

Imputation separates the missing data problem from the analysis of scientific interest. Consequently, the completed data can be analyzed as if it had been fully observed, without added complexity in the analysis of scientific interest.

This tutorial illustrates the imputation of incomplete multilevel data with the statistical programming language *R*. We aim to show how imputation can yield less biased estimates from incomplete clustered data. We provide practical guidelines and code snippets for different missing data situations, including non-ignorable missingness mechanisms. For reasons of brevity, we focus primarily on multilevel imputation using chained equations with the popular *R* package *mice*, in combination with other *R* packages which are used for applications and visualizations.

The case study datasets cover typical data structures from the social and biomedical sciences. These include an example of clustering in individual patient data meta-analyses and a ‘missing not at random’ missingness mechanism.

**Relevance to conference theme**

Missing data is ubiquitous in the human data sciences, such as the social and medical fields, where multilevel imputation methods are commonly applied. Our work showcases how recent developments in imputation methodology can be leveraged by multilevel experts.

**Keywords (max. 3)**

Missing data, imputation, R.