36

How many (multiple) imputations do I need for an important analysis?

Tim P. Morris[1], James R. Carpenter[1,2], Andrew Atkinson[3]

[1]MRC Clinical Trials Unit at UCL, London, United Kingdom. [2]LSHTM Department of Medical Statistics, London, United Kingdom. [3]Washington University in St. Louis, St. Louis, USA

Tim P. Morris
Please provide a brief biography for the Presenting author(s)
Tim is a principal research fellow based at the MRC Clinical Trials Unit at UCL. His research focuses on the development, evaluation and understanding of statistical methods. His research interests include simulation studies, handling missing data, estimands, causal inference, sensitivity analysis, covariate adjustment, and IPD meta-analysis. He is an author of the 2024 book 'Multiple Imputation and its Application'.

James R. Carpenter
Please provide a brief biography for the Presenting author(s)
James is a Professor of Medical Statistics at the MRC Clinical Trials Unit at UCL (50%), and at London School of Hygiene and Tropical Medicine (50%). His research interests include missing data (he is the first author of Multiple Imputation and its Application), estimands, statistical methods for meta-analysis, and neurodegenerative diseases.

Andrew Atkinson
Please provide a brief biography for the Presenting author(s)
Andrew is a biostatistician and Assistant Professor of Medicine at Washington University School of Medicine in St. Louis, with interests in methods for handling missing data, novel clinical trial designs, trial emulation and Bayesian modelling.

Single topic, multi-speaker session, Workshop or Single presentation submission

A single presentation/poster

Single presentation or poster submission

The validity, versatility and pitfalls of multiple imputation to handle missing data are well established. When multiple imputation is to be used, a common planning question is, 'How many imputations do I need?'

There are several valid answers, ranging from 'at least 2' to 'as many as feasible'. I will describe a principled and general two-step plugin method that aims to control so-called 'Monte Carlo error' (the variation in our inference due to a finite number of imputations). For important analyses, such as analysis of clinical trials, achieving a Monte Carlo error below some threshold should be our aim. The two-step plug-in approach explicitly targets this for one or more statistics of interest; for example, a point estimate, variance estimate, p-value or confidence limit. I will describe the approach and give some results of a simulation study evaluating the approach. This method provides reassurance that the results of importance analyses are also stable.