Statistics in Practice session, IBC Florence

Title: Practical approaches to the analysis of partially observed data

Presenters: James Carpenter[1,2] and Rod Little[3]

[1]Department of Medical Statistics, London School of Hygiene & Tropical Medicine

[2]Medical Research Council Clinical Trials Unit, Kingsway, London

[3] Department of Biostatistics, University of Michigan, Ann Arbor MI

Emails: james.Carpenter@lshtm.ac.uk   rlittle@umich.edu1

Extended abstract:

This session, consisting of two 105-minute presentations, discusses the issues raised by missing data and practical approaches for analyzing the resulting partially observed data. We review a range of methodological approaches, with a focus on multiple imputation. Our aim is that participants gain the knowledge and confidence to apply these techniques in their own work.

We consider issues raised by both randomised clinical trials (where missing data occur mostly in the dependent variables), and observational data (where missing data occur in both dependent and independent variables).

The sessions cover the following material:

Session 1:

- Introduction

    A non-trivial proportion of missing data introduces ambiguity into the analysis and resulting inference, because the reasons for values being unobserved are ultimately unknown. All analyses, and resulting inferences, therefore rest on untestable assumptions. In applications, it is therefore important to be clear about these assumptions, and then apply appropriate statistical methods.

    With this in mind, we begin by reviewing Rubin's typology for missing data mechanisms and use this to explore the assumptions underlying analyses restricted to complete records.

- Multiple Imputation

    We give an intuitive introduction to multiple imputation, and describe and contrast algorithms based on joint modelling and full conditional specification. Issues that arise when using these algorithms, and the capabilities, pros and cons of some available software are discussed. The potential and pitfalls of multiple imputation are illustrated with examples.

- Maximum likelihood and inverse probability weighting

We discuss maximum likelihood estimation with missing data, and outline the EM algorithm, and approaches derived from this. Inverse probability weighting and extensions to augmented inverse probability weighting are introduced and the attractions of these approaches vis-a-vis multiple imputation discussed.

- Non-linear relationships

    Choosing an appropriate imputation model is the key practical task for the analyst. The issues raised by non-linear relationships and interactions in the substantive model are presented, together with recent proposals for addressing them.

Session 2:

- Measurement error
    The multiple applications of multiple imputation make it very attractive. We consider the utility of multiple imputation for handling measurement error, and appropriate versions of Rubin's rules in this context.

- Survey weights
    Appropriate handling of survey weights with multiple imputation has been the subject of some discussion in the literature; we outline the key issues and how they can be addressed in practice.

- Hierarchical data
    Many datasets are hierarchical, potentially with missing values at all levels of the hierarchy. We discuss and illustrate appropriate imputation strategies and software in this context.

- Sensitivity analysis
    In applications, sensitivity analysis—where we explore the robustness of inferences to a range of plausible assumptions about the missing data mechanism—is often necessary. We contrast selection and pattern mixture models and how they may be approached using multiple imputation.

- Applications of sensitivity analysis
    We illustrate the use sensitivity analysis in clinical trial and observational settings.

We conclude with a general discussion and consider future research directions.

<u>References</u>

Little, R J A and Rubin, D B (2002) *Statistical analysis with missing data (second edition)* Chichester: Wiley

Carpenter, J R and Kenward, M K (2013) *Multiple imputation and its application.* Chichester: Wiley