

# Some issues in dealing with missing data

Harvey Goldstein, University of Bristol

# Multiple imputation – some limitations

- Requirement for ‘congeniality’ strictly rules out interactions in MOI.
- MAR assumption often not realistic – MNAR more likely
  - Research has focussed largely on MCAR, MAR models
  - How far can we use auxiliary data to approximate MAR?
  - In research studies what balance of effort should go into acquiring good auxiliary data vs improving response rates?
  - How can we move to incorporating proper missing data methods into standard software, and especially diagnostic procedures for exploring patterns of missingness?
  - Are there rules of thumb that can be invoked as to when we need to get sophisticated rather than for example, using casewise deletion?

# New methodology

- Joint (Bayesian) modelling for imputation and MOI (IMOI) solves problem of interactions and provides sounder inferences.
- Procedures for handling (survey) weights becoming available.
- For multilevel data we can have missingness in covariates at several levels (including aggregated variables) and this will be a fruitful area for further research and software development.
- How do we convey to data analysts the importance of sound missing data analysis?

# References to recent work

- Goldstein, H., Harron, K., and Wade, A. (2012). The analysis of record linked data using multiple imputation with data value priors. *Statistics in medicine*, 31, \*, DOI: 10.1002/sim.5508
- Goldstein, H., Carpenter, J. R. and Browne, W. J. (2014), Fitting multilevel multivariate models with missing data in responses and covariates that may include interactions and non-linear terms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 177(2), 553-564 doi: 10.1111/rssa.12022
- Goldstein, H., Carpenter, J., and Kenward, M. (2018). Bayesian models for weighted data with missing values: a bootstrap approach. *J. Royal Statistical Society, series C.*, 64, 4, 1071-1081. DOI: 10.1111/rssc.12259