Simultaneous treatment of measurement error and missing data in HIV treatment research using overimputation

18th International Workshop on HIV Observational Databases

Sitges, Spain

27 March 2014

Michael Schomaker, Sara Hogger, Leigh Johnson, Till Bärnighausen, Chris Hoffmann, Christian Heumann

for the IeDEA Southern Africa Cohort Collaboration

Background

- It is well-known that both CD4 and viral load in HIV infected persons are *measured with error*, both due to physiological/biological variation and technical constraints.
- *Missing* or unavailable laboratory *data* (at baseline) is common, especially in resource limited settings.
- While the problem of (ignorable) missing data can nowadays be handled easily with multiple imputation or IPW, methods addressing measurement error are often complicated to implement and restricted to a particular setting or model.
- Multiple overimputation, recently suggested in the area of political sciences (Blackwell et al., 2012), can deal with both missing data and data measured with error, can be easily implemented, and is applicable to most analytical questions.

Multiple Overimputation
(Blackwell et al., 2012)

1. a) multiply impute (say M times) missing values and
   b) multiply overimpute (replace) mismeasured values

   → treat mismeasured data as missing but use mismeasured values as prior information when imputing

   [utilize with `Amelia II` package in *R* (Honaker et al., 2011)]

2. Conduct any inference (e.g. Cox model, KM estimator, ...) on each overimputed set of data

3. Combine the M estimates related to the M overimputed sets of data according to standard MI combining rules

Measurement error model & imputation model

Classical measurement error model:

$$\text{observed } x_{ij}^* = \text{latent } x_{ij} + u_{ij}, \quad u_{ij} | x_{ij} \sim N(0, \sigma_{u_{ij}}^2)$$

In our situation, based on a literature review[1], we specify:

$$\begin{aligned} \ln \mathrm{CD4}_i^* &= \ln \mathrm{CD4}_i + u_i, & u_i | x_i \sim N(0, 0.26^2) \\ \log_{10} \mathrm{VL}_i^* &= \log_{10} \mathrm{VL}_i + u_i, & u_i | x_i \sim N(0, 0.255^2) \end{aligned}$$

Use Amelia II imputation algorithm[2], and add observation level priors to (multivariate normal) imputation model:

$$\begin{aligned} \ln \mathrm{CD4}_i &\sim N(\ln \mathrm{CD4}_i^*, 0.26^2) \\ \log_{10} \mathrm{VL}_i &\sim N(\log_{10} \mathrm{VL}_i^*, 0.255^2) \end{aligned}$$

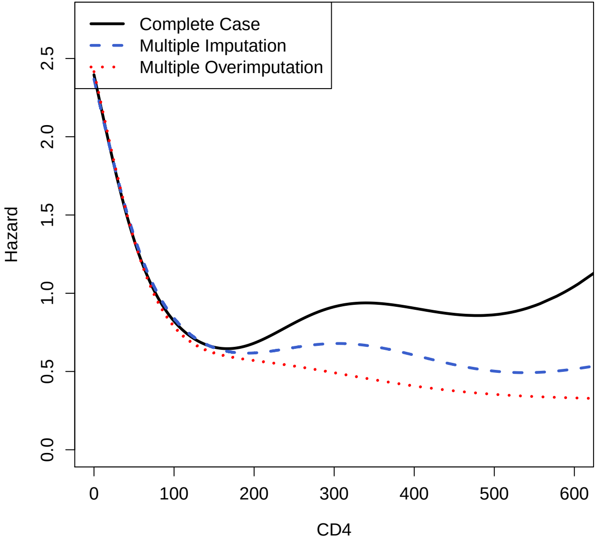
[1] For example, Lew et al. (1998) and Hoover et al. (1992) among others

[2] Honaker et al. (2011)

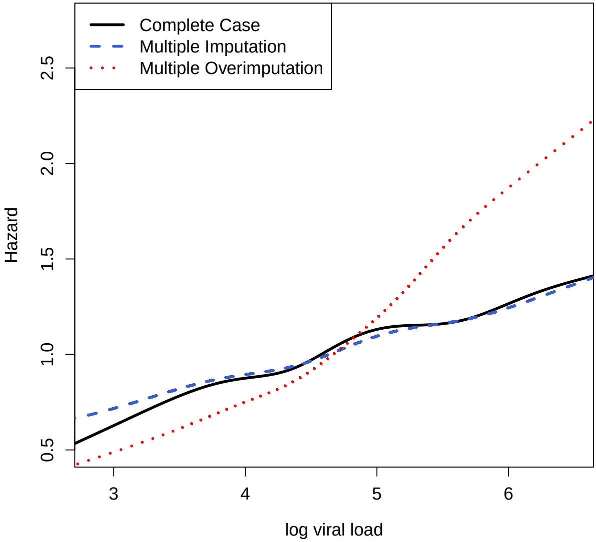
Data Analysis

- We use IeDEA-SA data of nearly 30,000 patients from 4 South African cohorts.
- We apply multiple overimputation:
  - imputation of missing baseline CD4 and VL values
  - overimputation of all measured baseline CD4 and VL values
  - all covariates (baseline, demographics, outcome) included in imputation model
- We estimate a Cox proportional hazards model to estimate the effect of baseline CD4 and baseline log viral load on the hazard of death, adjusted for year of treatment initiation, sex, cohort, and age.
- The effects of CD4 and log viral load are modelled non-linearly via p-splines.

Effect of CD4

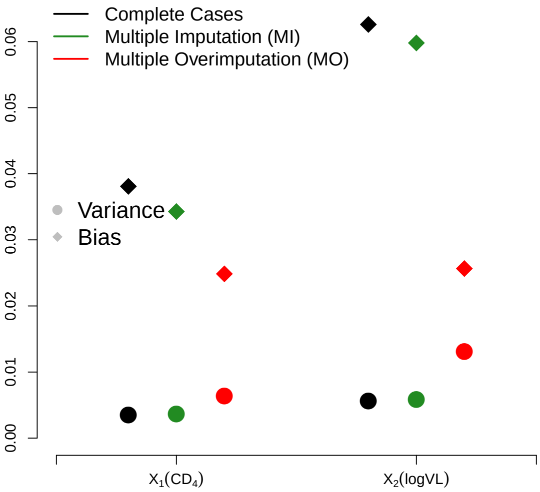


Effect of log viral load



Simulations

Simulated survival data with measurement error and missing data (MAR), 2500 repetitions, evaluate parameters of Cox models.



MO removes bias, but sometimes at the expense of a higher variance.

IeDEA Southern Africa
International epidemiologic Databases to Evaluate AIDS

Conclusions

- Multiple overimputation offers a convenient approach to account for both mismeasured and missing data and can be easily implemented.

- In our analysis, it is likely that standard multiple imputation or complete case analyses led to attenuated estimates, which could be corrected by means of multiple overimputation.

- Preliminary simulation study show a good performance of multiple overimputation, but more work needs to be done to fully understand limitations and implications of the approach.

Acknowledgements

- All patients and staff from participating sites.
- Renee de Waal
- Funders: This work was supported by the US National Institute of Allergy and Infectious Diseases (NIAID) through the International epidemiological Databases to Evaluate AIDS, Southern Africa (IeDEA-SA).

References

Blackwell, M., J. Honaker, and G. King (2012). Multiple overimputation: a unified approach to measurement error and missing data. *Working paper, submitted*. http://gking.harvard.edu/amelia.

Honaker, J., G. King, and M. Blackwell (2011). Amelia II: A program for missing data. *Journal of Statistical Software 45(7)*, 1–47.

Hoover, D., N. Graham, B. Chen, J. Taylor, J. Phair, S. Zhou, and A. Munoz (1992). Effect of CD4+ cell count measurement variability on staging HIV-1 infection. *Journal of Acquired Immune Deficiency Syndromes 5*, 794–802.

Lew, J., P. Reichelderfer, M. Fowler, J. Bremer, R. Carroll, and et al. (1998). Determinations of levels of human immunodeficiency virus type 1 RNA in plasma: reassessment of parameters affecting assay outcome. *Journal of Clinical Microbiology 36*, 1471–1479.