Multiple imputation for covariate measurement error

Jonathan W. Bartlett[1], Bianca L. De Stavola[1] and Chris Frost[1]

[1] Medical Statistics Unit, London School of Hygiene and Tropical Medicine, UK

It is well known that measurement error in the covariates of regression models causes bias in parameter estimates. Allowance for this type of error can be made when additional information regarding the measurement error process is available, typically from a replication or validation sub-study, and many estimation methods have been proposed, including method of moments, regression calibration, simulation extrapolation (SIMEX) and maximum likelihood (ML).

One approach is to view the problem of covariate measurement error as a missing data problem, and use multiple imputation (MI). This is the approach adopted by Cole et al [1], who demonstrated MI in the case in which for a subset of study subjects, both the 'true' covariate X and an error-prone measurement of it, W are observed (validation study). For this type of data, existing software for MI can be used to impute X for those subjects who do not belong to the validation sub-study.

However, there are several settings where it is not possible or feasible to observe the true covariate X without error (e.g. blood pressure), but where replicate error-prone measurements can instead be made. In this case, standard MI software cannot be used to impute X, because X is missing on all study subjects.

We illustrate how standard linear mixed models can be used to set up an appropriate imputation model when X is continuous and measured with error, from which imputations of X can be made. Analyses can then be performed using the imputed values of X, with no further adjustment for measurement error required. We examine the performance of the proposed implementation of MI for measurement error through simulations, and compare it to the performance of regression calibration and ML. We consider the implementation of MI for three types of outcome: a Normally distributed continuous outcome, a binary outcome, and a survival outcome subject to censoring.

Our simulation results show that for a continuous, Normally distributed outcome, our proposed implementation of MI gives identical point estimates to ML. We show that in the case where only a subset of subjects have replicate measurements (unbalanced case), MI and ML are more efficient than (unweighted) regression calibration. In this case, using MI is computationally easier and faster than finding the ML estimates via the expectation maximization (EM) algorithm. Lastly, we discuss situations in which using MI to deal with covariate measurement error may be advantageous relative to established methods such as regression calibration.

References

[1] Cole SR, Chu H, Greenland S (2006) Multiple-imputation for measurement error correction. International Journal of Epidemiology 35:1074-1081