

A comparison of methods for analysing incomplete longitudinal binary data

Rhian M. Daniel¹ and Michael G. Kenward¹

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

Clinical studies based on $t > 1$ repeated binary outcomes are common in medical research. If the data are incomplete, e.g. if some subjects drop out after $j < t$ visits, never to return, the missing data issue must be addressed.

One popular method for analysing correlated binary data is Generalised Estimating Equations (GEE). It is well-known that the validity of this method in its simplest form when the data are incomplete relies on the often implausible assumption of Missing Completely at Random (MCAR). However, there are conditions under which the MCAR assumption can be relaxed to Missing at Random (MAR). Variants of GEE have been proposed that allow more generally for MAR mechanisms, namely through the introduction of inverse probability weights in methods proposed by James Robins and colleagues. Others favour a combination of multiple imputation and GEE known as MI-GEE.

Many simulation studies in the literature have compared the performance of some or all of these methods; see, for example, [1] and [2]. Inconsistencies in some aspects of the results between different simulation studies suggest a need for a more theoretical approach to comparing these methods. This is what we attempt in this talk.

First, we show that a Markov structure for the full data is sufficient for the consistency of un-weighted GEE under MAR when the correlation structure is correctly specified. Using this as a basis, we quantify the bias that can be expected when the data are non-Markov and conclude that in many plausible longitudinal data structures, the bias is small. In the same setting, we also consider the relative precision of these estimators.

After confirming these theoretical results through some simulations, we compare the results obtained from applying the methods considered to data from a clinical trial comparing three drugs used in different combination therapies for type II diabetes patients.

References

- [1] Beunckens, C., Sotto, C. and Molenberghs, G. (2008) A simulation study comparing weighted estimating equations with multiple imputation based estimating equations. *Computational Statistics & Data Analysis* **52**:1533–1548
- [2] Preisser, J. S., Lohman, K. K. and Rathouz, P. J. (2002) Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine* **21**:3035–3054