*Statistical topics: Longitudinal data analysis*

*Application areas: Medical research*

**Selecting outcome measures and statistical models in neurological clinical trials that measure functional outcomes**

**Tea Uggen1**, Dr. Tapan Rai1, Dr. Erin Godecke2

*1 School of Mathematical and Physical Sciences, The University of Technology Sydney, Australia*

*2 School of Psychology and Social Science, Edith Cowan University, Australia*

In clinical trials that assess recovery or management of a neurological condition, it is common to use a variety of outcome variables to assess functional recovery. For example, a clinical trial may involve the assessment of aphasia, a communication disorder that is caused by a stroke or traumatic brain injury. Aphasia recovery is assessed using the Western Aphasia Battery Aphasia Quotient (AQ) and Discourse Analysis (DA), in addition to other measures. A number of alternative measures have also been derived from these, to facilitate statistical analysis.

While determining the most appropriate measure to use is a clinical issue, statistical considerations can inform the use of the most appropriate models to assess these measures. However, the problem of determining the most appropriate variant of an outcome measure to model is non-trivial. Although Information Criteria (eg. AIC, BIC) are commonly used in model selection, these criteria focus on the selection of predictors rather than on the selection of outcome variables. In this work, we demonstrate that Information Criteria cannot be used to select the best variant of an outcome variable since their values are scale-dependant and can change dramatically when a variable is transformed.

A more appropriate method to achieve this involves the examination of model residual plots. However, residual plots are not commonly used to assess the fits of longitudinal models such as generalized estimating equations (GEE) or linear mixed models, which are used to analyse functional recovery. In this work, we demonstrate the use of model diagnostic plots that were proposed by Park and Lee (2004) for GEE models, to determine the most appropriate variant of discourse analysis (a measure of aphasia recovery after a stroke). While our focus is on selecting the best model to assess aphasia recovery using discourse analysis, the method is general enough to be applied to the selection of the appropriate variant of an outcome measure in any clinical trial that uses GEE models to assess functional recovery.