A CLUSTER-DELETION TEST FOR LEVERAGE: AN INTUITIVE, USER-FRIENDLY GEE DIAGNOSTIC

M. Wilchesky $^{\dagger},$ R. Platt, R. Tamblyn

McGill University, Montreal, Canada

[†]E-mail: *machelle.wilchesky@mail.mcgill.ca*

Several methods exist for detecting outliers in regression models. However, to our knowledge, only one reference published by Preisser and Qaqish (1996) refers specifically to diagnostic statistics for evaluating outliers in general estimating equations (GEEs). We propose a new cluster-deletion technique where the leverage accruing to each individual cluster is measured in terms of its impact on the working correlation matrix. As such, clusters with a disproportionate influence on the estimated GEE parameters can be identified. This method will work for the exchangeable correlation assumption as it compares the value of the off-diagonal element, or intra-class correlation, ρ .

In a model with k independent clusters, the GEE model is run k times, each time removing a different cluster from the data. The resulting distribution of k off-diagonal ρ values produced is then normalized by subtracting the mean and then dividing by the standard deviations.

In an illustration using data from a study examining the predictors of prescription nonadherence in an elderly Canadian population with hypertension, clusters identified with higher cluster-deletion diagnostic z-scores were associated with higher percent changes in both the parameter estimates and standard errors. These results are contrasted and compared with estimates produced using the technique published by Preisser and Qaqish.