Analysis of cluster randomization trials for binary data

Jörg Kaufmann

Independent Statistical Consultant, Berlin, Germany

Abstract

Clustered data occur frequently in many fields of application. Toxicological experiments often involve animal litters as experimental units. Pregnant female animals are assigned to a control or one of several dosed groups. The responses for toxicity endpoints (effects) may be either continuous, e.g. fetal body weights, or dichotomous, e.g. presence or absence of a fetal anomalies. In diagnostic imaging studies, e.g. contrast enhanced angiography, particular vessel segments contributes information to the diagnosis.
The unit of inference in cluster randomization trials may be directed either at the cluster level or at the level of observational unit. Challenges in identifying and distinguished the unit of inference the units of analysis are not unique to cluster trials. Cluster–level analyses are most obviously appropriate when the primary question of interest focus more on the randomized unit as a whole than on the individual subjects or vessel segments.
The number of subjects required per treatment using standard sample size formula has to be multiplied under cluster randomization by a design effect term or variance inflation factor given by $\mathrm{IF} = 1 + (m - 1)\,\rho$, where m is the average cluster size and ρ is the prior estimate of the intracluster correlation coefficient.
One individual–level approach for testing the difference between event rates is to empirically adjust the standard chi–square statistic test for the clustering of responses within a experimental randomized unit. The relatively simple method is intuitively attractive since it yields the standard Pearson chi–square statistic if the estimated intracluster coefficient ρ is zero, and is easily extended to the comparison of more than two groups. It is also imposes no distributions on the responses within cluster. The principal aim of many clusters designs is to compare the proportion of units in different treatment groups. I will give examples of sample size estimation, analysis of binary outcome for completely randomized and matched-pairs design and compare test results between cluster–level and individual–level procedures.