MP53 Performance of mixed effects models for partially clustered trials

Lange K.*[1], Sullivan T.[2], Kasza J.[3], Yelland L.[2]
[1]The University of Adelaide ~ Adelaide ~ Australia, [2]South Australian Health and Medical Research Institute ~ Adelaide ~ Australia, [3]Monash University ~ Melbourne ~ Australia

Methods for designing and analysing cluster randomised trials are well established. However, many clinical trials involve partially clustered data, where only some observations belong to a cluster. For example, neonatal trials may include infants from single or multiple births, while ophthalmology trial participants may need treatment for one or both eyes. Recently, we defined four types of partially clustered trial designs characterised by whether the clustering occurs pre- or post-randomisation and the method of randomisation for clustered observations [1]. However, the performance of analysis methods for such trials, including mixed effects models and generalised estimating equations (GEEs), have received limited attention and sample size formulas are only available for GEEs [2]. The aims of this study were to assess (1) the performance of mixed models versus GEEs for analysis of partially clustered trials, and (2) whether existing sample size formulas based on GEEs provide appropriate power for analysis via mixed models. A simulation study was conducted in R to evaluate the performance of mixed models versus GEEs for estimating the effects of treatment on a continuous outcome. We considered a maximum cluster size of 2 and simulated datasets with the sample size required to achieve 80% power according to GEE- based formulas. Simulation parameters were chosen to reflect the range of scenarios observed in practice (effect size 0–0.8, ICC 0.1–0.9, proportion of paired observations 0.015–0.7, pairs randomised using cluster, individual or balanced randomisation). Datasets were analysed using mixed effects models and GEEs with an independence or exchangeable working correlation structure. GEEs generally performed well with some exceptions when the ICC was high with individual or balanced randomisation. Performance of the mixed model was typically comparable to GEEs, though non-convergence and under-coverage occurred (maximum non-convergence rate 11%, minimum coverage rate 75%) in more extreme settings (e.g. few pairs, high ICCs). Calculating the target sample size using the exchangeable correlation GEE resulted in approximately 80–85% power for the mixed model across all scenarios. In many partially clustered trial settings, both GEEs and mixed effects models perform well. Existing sample size formulas based on GEEs may be appropriate for analysis via mixed models.

[1] K.M. Lange, J. Kasza, T.R. Sullivan, L.N. Yelland. Clinical Trials, 20(2), 2023, pp 99-110
[2] L.N. Yelland, T.R. Sullivan, D.J. Price, K.J. Lee. Statistics in Medicine, 36(8), 2017, pp 1227–1239

Poster Sessions

MP54 10-Years changes in lung function of cystic fibrosis patients in europe: different statistical methods at work

Orenti A.*[1], Adamoli A.[1], Kerem E.[2], Hatziagorou E.[3], Zolin A.[1], Ambrogi F.[1]
[1]University of Milan, Department of Clinical Sciences and Community Health, Laboratory of Medical Statistics, Biometry and Epidemiology "G. A. Maccacaro" ~ Milano ~ Italy, [2]Hadassah University Medical Centre, Hebrew University Hadassah Medical School, Department of Paediatrics and Centre for Cystic Fibrosis ~ Jerusalem ~ Israel, [3]Aristotle University of Thessaloniki, Paediatric Pulmonology and Cystic Fibrosis Unit, 3rd Paediatric Department, Hippokration Hospital ~ Thessaloniki ~ Greece

Cystic fibrosis (CF) is the most common severe autosomal recessive disease in Europe, with pulmonary insufficiency as the main cause of death. For prognosis, forced expiratory volume in 1 second percent of predicted (FEV1pp), is regarded as the best generally available measure for assessing CF lung disease. Since FEV1pp has a slightly asymmetric distribution, it is often summarized using median and quartiles. However, when fitting regression models, the results are usually provided in terms of means. The aim of the current study is to explore changes in FEV1pp during the last decade, comparing results obtained with different statistical regression methods including random effects. To estimate the difference in FEV1pp values over 2011 and 2021, data of 18756 people with CF, homozygote for F508del mutation and included in the European Cystic Fibrosis Society Patient Registry, are used. Three regression models including a random effect for patients and with FEV1pp as response variable are fitted using R software: the classical generalized estimating equations (GEE) model with Gaussian family, a linear quantile mixed model (LQMM) [1], a Generalized Additive Models for Location, Scale and Shape (GAMLSS) [2] with Normal family distribution. Two different setting are explored: in the first one the year of follow-up is included as a continuous variable, in the second one it is included using dummy variables. The results of the different models are comparable in terms of coefficient estimates. GAMLSS provides the narrower confidence interval then GEE and LQMM when year is included as a continuous variable, LQMM gives the narrower CI then GEE and GAMLSS when year is included as dummy variables. The main problem in fitting models in R software on our big dataset, is the long computational time. To obtain coefficient estimates and standard errors: 50 minutes for GAMLSS and almost 6 hours for LQMM. In conclusion, these models need to be additionally compared in detail for diagnostic measures,. Further research is needed to fulfill the unmet need of providing robust regression coefficient estimates on mixed effects models on big datasets, also simulation studies mimic real world practice are necessary.

[1] M. Geraci, Linear quantile mixed models: The lqmm package for Laplace quantile regression. Journal of Statistical Software, 57(13), 2014,1-29.
[2] D.M. Stasinopoulos, R.A. Rigby, Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software, 23(7), 2007, 1-46.

MP55 Structural pathway analysis of longitudinal multinomial phenotypes

Kamruzzaman M.[1], Park T.*[2]
[1]Jagannath University ~ Dhaka ~ Bangladesh, [2]Seoul National University ~ Seoul ~ Korea, Republic of

Several statistical methods for pathway analysis have been developed to test the association between pathways and phenotypes of interest. Since pathways are highly correlated, thus a hierarchical structural component model (HisCoM) was developed to analyze all pathways in a single model and take into consideration their correlation. HisCoM was originally developed to analyze a single phenotype using only one measurement per individual. Later, it was extended to analyze multiple phenotypes (HisCoM-multi) and longitudinal phenotypes (HisCoM-GEE). These methods have been used to analyze continuous, counts, and binary phenotypes from cross-sectional, clustered, and longitudinal studies. In this study, we propose a hierarchical structural component model for pathway analysis of longitudinal multinomial phenotypes (HisCoM-RCateg). HisCoM-RCateg is proposed by combining the hierarchical structural component model and generalized estimating equations for correlated multinomial phenotypes. HisCoM-RCateg accounts for the biological hierarchy of all biomarkers and pathways into a single model. In the simulation, the proposed HisCoM-RCateg appeared to have high power than other existing methods and effectively controlled type I error for longitudinal multinomial phenotypes. To demonstrate the performance, we also applied HiscoM-RCateg to two distinct types of longitudinal omics data, namely the metabolite dataset and the metagenome dataset. HisCoM-RCateg has an advantage of taking into account the true biological hierarchical structure directly into the statistical model.