Illustrative GEE Analysis of Cluster Randomized Crossover (CRXO) Trials

Fan Li

Trial Data with Continuous Outcomes

We first read in the simulated trial data set with continuous individual outcomes.

```
dir<-"D:/Research/CRT Methodology/CRXOSampleSize/Latex/Submission/R Code"

setwd(dir)

simdata_cont<-read.csv("simdata_cont.csv", header = TRUE)
```

The first 6 rows of the data set looks like the following:

```
head(simdata_cont)
```

```
##              y ind cluster period period1 period2 treatment
## 1 -0.612354568   1       1      1       1       0         1
## 2 -0.716552226   2       1      1       1       0         1
## 3  0.009402725   3       1      1       1       0         1
## 4  1.482727258   4       1      1       1       0         1
## 5  1.009020225   5       1      1       1       0         1
## 6  0.908986321   6       1      1       1       0         1
```

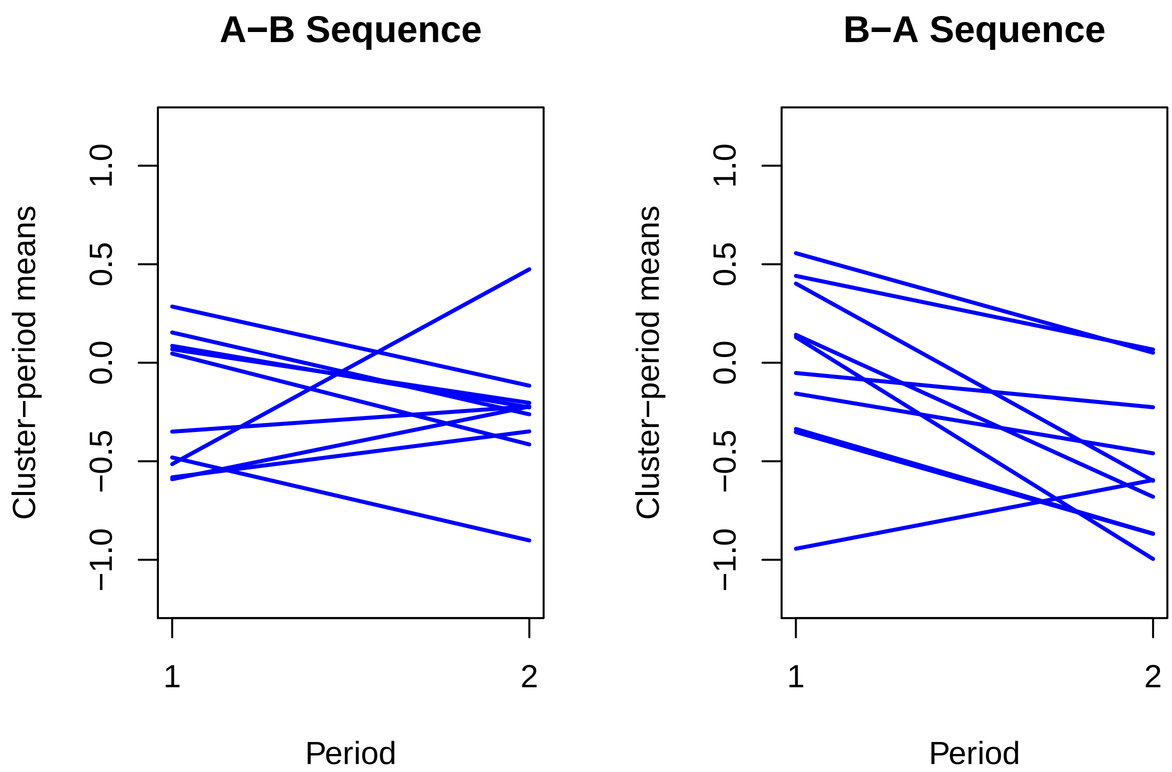
From left to right, the columns of data are outcome (in long format), individual id, cluster id, period id, indicator for period 1, indicator for period 2 and indicator for treatment. Note that this is a simulated two-treatment two-period, cross-sectional CRXO trial with 20 clusters, 65 individuals per cluster (cluster-period size). We will extract the following key elements from the data set.

```
# outcomes
y<-as.numeric(simdata_cont$y)

# cluster identifier
id<-as.numeric(simdata_cont$cluster)

# period identifier
period<-as.numeric(simdata_cont$period)

# marginal mean design matrix
X<-simdata_cont[,c("period1","period2","treatment")]
X<-as.matrix(X)

# treatment indicator
trt<-X[,"treatment"]

# number of clusters
n<-length(unique(id))

# number of periods
t<-dim(X)[2]-1
```

```
# cluster size (across all periods)
clsize<-as.numeric(table(id))

# cluster-period size (per cluster; balanced)
m<-clsize/t
```

By summarizing the cluster-period means by the following plot, we conjecture that there is a gently decreasing period effect over time, and the treatment (denote A, B as treatment and control) appears to reduce the magnitude of the outcome.

```
# Cluster-period means
clp_mu<-tapply(y,list(id,period),FUN=mean)
```



To perform the GEE and MAEE analysis of the trial data, we need to obtain the design matrix for the correlation parameters as follows.

```
# Create (large) design matrix for correlations
CREATEZ<-function(n,m,t){
  # correlation position indicators
  alpha0_pos<-1
  alpha1_pos<-2
  zrow<-diag(2)
  Z<-NULL

  for(i in 1:n){
    mi<-m[i]
    bm1<-(1-alpha0_pos)*diag(t*mi)
```

```
    bm2<-(alpha0_pos-alpha1_pos)*kronecker(diag(t),matrix(1,mi,mi))
    bm3<-alpha1_pos*matrix(1,t*mi,t*mi)
    POS<-bm1+bm2+bm3

    for(j in 1:(t*mi-1)){
      for(k in (j+1):(t*mi)){
        Z<-rbind(Z,zrow[POS[j,k],])
      }
    }
    # print(i)
  }
  return(Z)
}

# large matrix (may take a minute to run)
Z<-CREATEZ(n,m,t)
```

We confirm the exploratory analysis of the trial data by fitting the GEE and MAEE using the following code. Detailed descriptions of the input arguments are available in the `contMAEE.R` program. Following the notations in Li, Forbes, Turner and Preisser, we use the marginal mean model

$$\mu_{ijk} = \tau_j + \delta X_{ij},$$

where the link is the identity function (canonical link), τ_j is the jth period effect, X_{ij} is the treatment indicator of cluster i in period j, δ is the marginal intervention effect. Further, the nested exchangeable correlation structure is parameterized with two correlation parameters (α_0, α_1).

```
# Source the function
source("contMAEE.R")

# Implement the function
contMAEE(y=y,X=X,id=id,n=clsize,Z=Z,maxiter=25,epsilon=0.001,printrange="NO",
         shrink="ALPHA",makevone="NO")
```

```
## Loading required package: MASS

## GEE for correlated Gaussian data
##  Number of Clusters: 20
##  Maximum Cluster Size: 130
##  Minimum Cluster Size: 130
##  Number of Iterations: 5
## Results for marginal mean parameters
##      Beta     Estimate  MB-stderr BC0-stderr BC1-stderr BC2-stderr
## [1,]    0  0.008625779 0.09746112 0.09888618 0.10283157  0.1069617
## [2,]    1 -0.269863609 0.09746112 0.09589852 0.09987672  0.1040587
## [3,]    2 -0.221990814 0.10038730 0.09588854 0.10107539  0.1065428
##      BC3-stderr
## [1,] 0.10256656
## [2,] 0.09942105
## [3,] 0.09967665
##
## Results for correlation parameters
##      Alpha   Estimate BC0-stderr BC1-stderr BC2-stderr BC3-stderr
## [1,]     0 0.11945781 0.07452745 0.07728893 0.08017241 0.07641222
## [2,]     1 0.03697958 0.02683795 0.02756458 0.02832107 0.02755009
```

The treatment effect estimate is -0.22, which is slightly larger but close to the truth, -0.25, used in the data generation. A gently decreasing period effect is reflected in the parameter estimates. The estimates for within-period correlation α_0 and inter-period correlation α_1 are close to the true values, 0.1 and 0.05. For the class of bias-corrected sandwich variances, we observe that BC0<BC1<BC2 and BC1≈BC3. Note that this is an illustrative analysis of only one data set with a limited number of clusters.

Trial Data with Binary Outcomes

We read in the simulated trial data set with binary individual outcomes.

```
simdata_bin<-read.csv("simdata_bin.csv", header = TRUE)
```

The first 6 rows of the data set looks like the following:

```
head(simdata_bin)
```

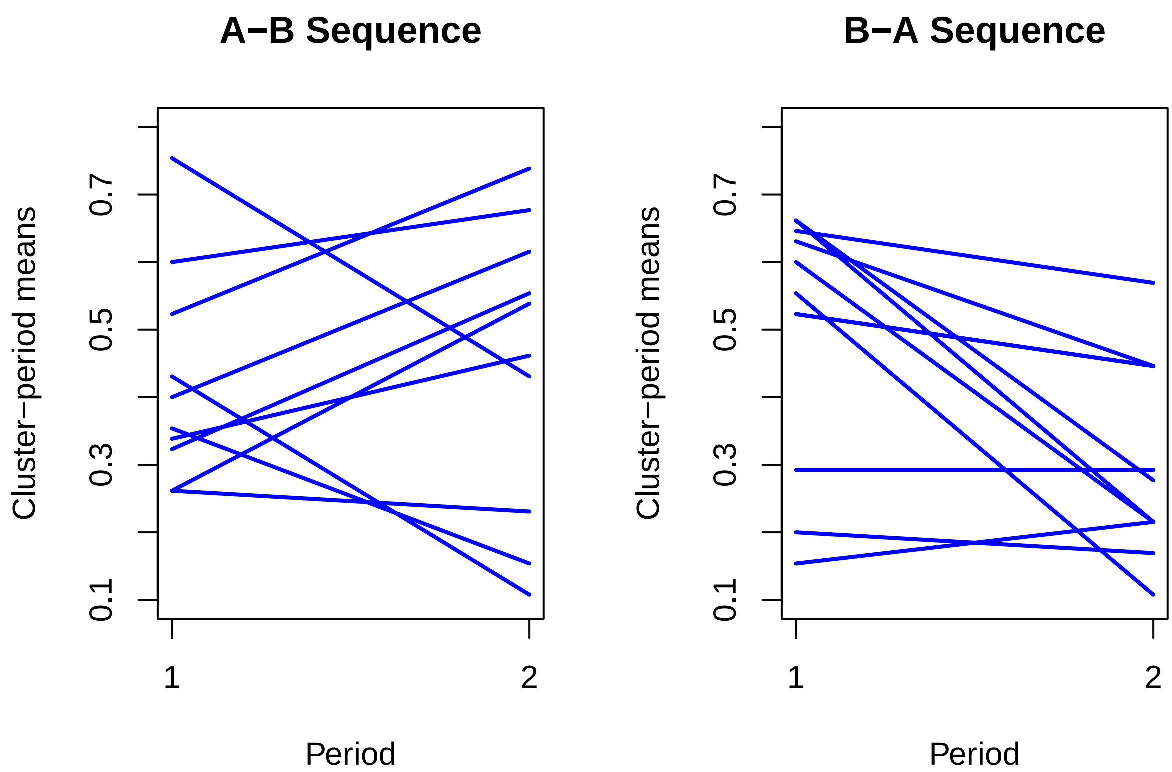
```
##   y ind cluster period period1 period2 treatment
## 1 0   1       1      1       1       0         1
## 2 0   2       1      1       1       0         1
## 3 0   3       1      1       1       0         1
## 4 0   4       1      1       1       0         1
## 5 0   5       1      1       1       0         1
## 6 0   6       1      1       1       0         1
```

We will extract the following key elements from the trial data set as before.

```
# outcomes
y<-as.numeric(simdata_bin$y)

# cluster identifier
id<-as.numeric(simdata_bin$cluster)

# period identifier
period<-as.numeric(simdata_bin$period)

# marginal mean design matrix
X<-simdata_cont[,c("period1","period2","treatment")]
X<-as.matrix(X)

# treatment indicator
trt<-X[,"treatment"]

# number of clusters
n<-length(unique(id))

# number of periods
t<-dim(X)[2]-1

# cluster size (across all periods)
clsize<-as.numeric(table(id))

# cluster-period size (per cluster; balanced)
m<-clsize/t
```

By summarizing the cluster-period rates by the following plot, we conjecture that there is a gently decreasing period effect over time, and the treatment appears to be associated with decreased rates.

```r
# Cluster-period means
clp_mu<-tapply(y,list(id,period),FUN=mean)
```



Since this trial data possess the same structure as the previous one (20 clusters, 65 individuals per cluster-period), we could use the same design matrix for estimating correlation parameters. We confirm the exploratory analysis of the trial data by fitting the GEE and MAEE in the following analysis. We use the marginal mean model

$$\text{logit}(\mu_{ijk}) = \tau_j + \delta X_{ij},$$

where the link is the logistic function (canonical link), τ_j is the jth period effect, X_{ij} is the treatment indicator of cluster i in period j, δ is the marginal intervention effect on the log odds ratio scale. Again, the nested exchangeable correlation structure is parameterized with two correlation parameters (α_0, α_1).

```r
# Source the function
source("binMAEE.R")

# Implement the function
binMAEE(y=y,X=X,id=id,n=clsize,Z=Z,maxiter=25,epsilon=0.001,printrange="NO",
        shrink="ALPHA",makevone="NO")
```

```
## GEE for correlated binary data
##  Number of Clusters: 20
##  Maximum Cluster Size: 130
##  Minimum Cluster Size: 130
##  Number of Iterations: 6
## Results for marginal mean parameters
##      Beta    Estimate MB-stderr BC0-stderr BC1-stderr BC2-stderr
## [1,]    0  0.06527516 0.1943056  0.1935671  0.2013592  0.2095183
```

```
## [2,]    1 -0.29367662 0.1966496  0.2162917  0.2251987  0.2345260
## [3,]    2 -0.46636922 0.2031504  0.1897604  0.2000806  0.2109626
##      BC3-stderr
## [1,]  0.2009637
## [2,]  0.2247018
## [3,]  0.1972477
##
## Results for correlation parameters
##      Alpha   Estimate BC0-stderr BC1-stderr BC2-stderr BC3-stderr
## [1,]     0 0.12447295 0.02134092 0.02191888 0.02251259 0.02188142
## [2,]     1 0.03997813 0.02138419 0.02194831 0.02252732 0.02193774
```

We observe the marginal treatment effect in the odds ratio scale to be $\exp(\hat{\delta}) \approx 0.63$, which is close to the true value, 0.6, used in the data generation. A gently decreasing period effect is reflected in the parameter estimates. The estimates for within-period correlation α_0 and inter-period correlation α_1 are close to the true values, 0.1 and 0.05. We note that this is an illustrative analysis of only one data set with a limited number of clusters.