MMRM: Macro for Selecting Best Covariance Structure with the Method of Interest

Linga Reddy Baddam, inVentiv Health Clinical, Hyderabad, India
Sudarshan Reddy Shabadu, inVentiv Health Clinical, Hyderabad, India
Chunxue Shi, inVentiv Health Clinical, Maryland, USA

ABSTRACT

In clinical trial analysis, while handling longitudinal continuous data, there are very often cases that the Mixed Model Repeated Measures (MMRM) tool is used to deal with the continuous endpoints when an outcome is collected multiple times. It is usually up to the statistician to specify the criterion for identifying the best covariance structure against the chosen model among all possible covariance structures arranged in particular order of interest, or in the way it has been specified in Statistical Analysis Plan and/or Protocol. In order to achieve this with the use of SAS, clinical programmers have to develop a program which dynamically checks the each covariance structure by order of interest, or in the way it has been specified in Statistical Analysis Plan and/or Protocol, and to produce the estimates and main effects information based on the best covariance structure selected for the given model, The features of the %MMRM macro are explicitly discussed further in detail in this paper.

INTRODUCTION

Over the past few years, as insightful research has been increased to a greater extent in clinical trials to evaluate best possible statistical analysis of incomplete longitudinal data; there are always two challenges that the clinical industry faces:

Statistician's-view: What is the preferred statistical method for the available incomplete data, following Missing at Random pattern (MAR)?

Programmer's-view: How to convert the statistician's view into SAS code and present the justifiable result in clinical reports?

To examine statistician's-view further in much more detail, it is assumed that there are many covariance structures built of user interest in SAS repository. It is the statistician's call to penetrate deeper into data patterns and study conditions to select possible covariance structures for the R matrix, order and method of criterion.

THE MOST COMMONLY USED METHODS FOR FINDING APPROPRIATE COVARIANCE STRUCTURE

- **ORDER** - In the MMRM analysis, the within-patient errors will be modeled using an unstructured covariance matrix. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. If the model fails to converge, the covariance structures will be evaluated in the order given in Statistical Analysis Plan and/or Protocol until model convergence is met.
- **AIC** - In the MMRM analysis, the variance covariance structure which minimizes the Akaike's information criterion (AIC) from list of potential covariance structures given in Statistical Analysis Plan and/or Protocol is used for the estimation of the within-patient errors. The Kenward-Roger method is used to estimate denominator degrees of freedom.
- **AICc** - In the MMRM analysis, the variance covariance structure which minimizes the AICc from list of potential covariance structures given in Statistical Analysis Plan and/or Protocol is used for the estimation of the within-patient errors. The Kenward-Roger method is used to estimate denominator degrees of freedom.

  Note: AICc is AIC with a correction for finite sample sizes, and with a greater penalty for extra parameters. Therefore, Burnham & Anderson (2002) strongly recommend using AICc, rather than AIC, if *sample size* is small or *independent parameter number* is large.

- **BIC** - In the MMRM analysis, the variance covariance structure which minimizes the Bayesian information criterion (BIC) from list of potential covariance structures given in Statistical Analysis Plan and/or Protocol is used for the estimation of the within-patient errors. The Kenward-Roger method is used to estimate denominator degrees of freedom.

The following algorithm will convert the statisticians' view into programming steps.

1) Getting all the covariance structures of interest in bucket and queuing them up by priority
2) Selecting the method to find best covariance structure
3) Picking up each covariance structure from the queue and running against the model to check whether a convergence criterion is met or not.
4) Choosing the best covariance structure which converged using one of the above methods, and running the model with that covariance structure to obtain the estimates, fixed effects to present in the clinical report.

The following is the macro has been developed to implement the algorithm.

MACRO OVERVIEW

While the primary goal of this macro is to evaluate the best covariance structure based on method of interest while using a mixed model repeated measurement (MMRM) analysis for available incomplete data, the macro also includes an ORDER value for the sel_methd macro parameter, which allows the user to specify their choice of covariance structure.

The macro call is comprised of nine keyword parameters:

```
%MACRO mmrm(indata =,
            dep_var =,
            ind_var =,
            cls_var =,
            ddfm_methd =,
            rep_var =,
            subject =,
            cov_list =,
            lsmn_var =,
            sel_methd =,
            los =);
```

The %mmrm macro call is comprised of nine keyword parameters,

indata:

   Required - name of the SAS dataset containing the required information for MMRM model.

dep_var:

   Required - variable name of the outcome interested (dependent variable).

ind_var:

   Required - list of independent factors (variables) which you want test separated by space.

cls_var:

   Required - list of classification variables from the independent variables in the model separated by space. Classification variables can be either character or numeric.

ddfm_methd:

   Required - specifies the method for determining denominator degrees of freedom.

rep_var:

   Required - name of the repeated effect variable.

Subject:

   Required - name of the subject identification variable in the mixed model.

cov_list:

   Required - specifies the covariance structures to be evaluated for model convergence separated by hash #.

lsmn_var:

Required – specifies the fixed effect(s) from the model statement that involves the CLASS variables. You can specify multiple effects separated by space.

sel_methd:

Required - valid values = (ORDER, AIC, AICc and BIC)

- ORDER - evaluates the listed covariance structures one by one in the given order, and returns the results with the first covariance structure where the model converges.

- AIC - evaluates all of the listed covariance structures, and returns the results with the covariance structure which converges the model with the minimum AIC value (error or variance), it selects the first and foremost covariance structure in case of tie occurs among the converged covariance structures from the given order to parameter cov_list.

- AICc - evaluates all of the listed covariance structures, and returns the results with the covariance structure which converges the model with the minimum AICc value (error or variance) , it selects the first and foremost covariance structure in case of tie occurs among the converged covariance structures from the given order to parameter cov_list.

- BIC - evaluates all of the listed covariance structures, and returns the results with the covariance structure which converges the model with the minimum Bayesian information criterion (BIC) value (error or variance) , it selects the first and foremost covariance structure in case of tie occurs among the converged covariance structures from the given order to parameter cov_list.

los:

Required - specifies the alpha value to be used when constructing confidence interval for each of the LS-means.

```
%MACRO mmrm(indata =,
            dep_var =,
            ind_var =,
            cls_var =,
            ddfm_methd =,
            rep_var =,
            subject =,
            cov_list =,
            lsmn_var =,
            sel_methd =,
            los =);

  *Macro for checking convergence criteria;
  %MACRO cnvg_chk(cov =);
    *Deleting data sets cnvg_status and min_var if already exist;
    %IF %SYSFUNC(EXIST(cnvg_status)) %THEN %DO;
      PROC SQL;
        DROP TABLE cnvg_status;
        QUIT;
    %END;
    %IF %SYSFUNC(EXIST(min_var)) %THEN %DO;
      PROC SQL;
        DROP TABLE min_var;
        QUIT;
    %END;
    %IF %SYSFUNC(EXIST(%SCAN(%SCAN(&cov_list.,&i.,"#"),1,"("))) %THEN %DO;
      PROC SQL;
        DROP TABLE %SCAN(%SCAN(&cov_list.,&i.,"#"),1,"(");
        QUIT;
    %END;
    ODS LISTING CLOSE;
    PROC MIXED DATA = &indata.;
      CLASS &cls_var.;
```

3

```sas
      MODEL &dep_var. = &ind_var./HTYPE = 3 DDFM = &ddfm_methd.;
      REPEATED &rep_var./TYPE = &cov. SUB = &subject.;
      ODS OUTPUT CONVERGENCESTATUS = cnvg_status FITSTATISTICS = min_var;
    QUIT;
    ODS LISTING;
    %GLOBAL cng;
    %IF %SYSFUNC(EXIST(cnvg_status)) %THEN %DO;
      DATA _null_;
        SET cnvg_status;
        IF LOWCASE(reason) = "convergence criteria met." THEN
           CALL SYMPUT("cng", "yes");
        ELSE CALL SYMPUT("cng", "no");
      RUN;
    %END;
    %ELSE %LET cng = no;;
    %put cng = &cng.;
    %IF &cng. = yes and %UPCASE(&sel_methd) ^= ORDER %THEN %DO;
      %IF %SYSFUNC(EXIST(min_var)) %THEN %DO;
        DATA %SCAN(%SCAN(&cov_list., &i.,"#"), 1, "(")(KEEP = value cov);
          SET min_var(WHERE = (UPCASE(descr) =
                      "%UPCASE(&sel_methd.) (SMALLER IS BETTER)"));
          Cov = "%SCAN(&cov_list.,&i.,"#")";
        RUN;
      %END;
    %END;
  %MEND cnvg_chk;
  %GLOBAL fin_cov;
  *Counting number of covarince structures;
  %LET numcov = %SYSFUNC(COUNTW(&cov_list., '#'));
  %PUT numcov = &numcov.;
  %IF %UPCASE(&sel_methd) = ORDER %THEN %DO;
    *Identifying the covariance structure which coverged first from the given list of
covariance structures in order;
    %LET i = 1;
    %DO %UNTIL(&i. le &numcov. or %LOWCASE(&cng.) EQ yes);
      %cnvg_chk(cov = %SCAN(&cov_list.,&i.,"#"));
      %LET fin_cov = %SCAN(&cov_list.,&i.,"#");
      %LET i = %EVAL(&i.+1);
    %END;
    %PUT final covariance structure = &fin_cov.;
  %END;
  %ELSE  %IF  %UPCASE(&sel_methd.)  =  AIC  or  %UPCASE(&sel_methd)  =  BIC  or
%UPCASE(&sel_methd) = AICC %THEN %DO;
    *Selecting the covariance structure with minimum variance from the given list of
covariance structures;
    %DO i = 1 %TO &numcov.;
      %cnvg_chk(cov=%SCAN(&cov_list.,&i.,"#"));
      %LET covnm=%SCAN(&cov_list.,&i.,"#");
    %END;
    DATA mmrm;
      LENGTH cov $ 20;
      SET       %DO       i      =      1       %to       &numcov.;       %IF
%SYSFUNC(EXIST(%SCAN(%SCAN(&cov_list.,&i.,"#"),1,"("))) %THEN
      %SCAN(%SCAN(&cov_list., &i.,"#"),1,"("); %END;;
    RUN;
    PROC SQL NOPRINT;
    SELECT cov INTO: fin_cov FROM mmrm HAVING value = MIN(value) ;
    quit;
    %PUT final covariance structure = &fin_cov.;
  %END;
  %IF %NRSTR(&fin_cov.) ne %THEN %DO;
   *Taking the results using selected covariance structure above based given criteria;
    PROC MIXED DATA = &indata.;
```

```
          CLASS &cls_var.;
          MODEL &dep_var. = &ind_var./HTYPE = 3 DDFM = &ddfm_methd.;
          REPEATED &rep_var./TYPE = &fin_cov. SUB = &subject.;
          LSMEANS &lsmn_var./DIFF ALPHA = &los.;
          ODS OUTPUT LSMEANS = lsmean DIFFS = lsdiff TESTS3 = effects;
        QUIT;
    %END;
%MEND mmrm;
```

USING THE MACRO

Example

Suppose we have clinical trial data with a sample of 20 subjects randomly assigned to placebo and a new drug with a 1:1 randomization ratio. The outcome is a lab test measured at four time points. The first measurement is a pre-dose baseline assessment which is followed by three post-dose repeated measurements.

The following is the sample data for lab parameter:

```
DATA sample;
  INPUT subjid trt v0 v1 v2 v3;
  result=v0; visid=0; OUTPUT;
  result=v1; visid=1; OUTPUT;
  result=v2; visid=2; OUTPUT;
  result=v3; visid=3; OUTPUT;
  DROP v0-v3;
  CARDS;
 1 2 11.37 12.92   6.96  6.59
 2 1  9.65  3.21  10.94    .
 3 2 16.86 17.04  16.79 11.59
 4 2 13.99  4.51  12.2  13.93
 5 1  9.66  6.46  -1.2   9.09
 6 2  9.37  5.67  14.64  7.9
 7 1  4.63 16.66  11.14  5.68
 8 1  7.06 12.54  15.39 15.52
 9 2  5.44   7.3  10.8  10.88
10 1  6.51 17.0    5.16 14.22
11 2  8.97 17.52   3.38    .
12 1 12.15 12.97  10.32 20.46
13 2 10.75 14.37  12.89 17.18
14 2 10.43 10.32   8.17    .
15 1 13.68 16.97  14.28  9.84
16 2  9.49 19.84  17.07 14.05
17 1  8.24 14.65   5.19    .
18 1  1.39 12.73  11.25 10.23
19 1 10.82  2.61   9.18  7.19
20 2 10.45 17.39  17.08    .
;
run;
/* trt 1 = New drug, trt 2 = Placebo */

**** DERIVE CHANGE AND BASELINE VARIABLES;
PROC SQL;
  CREATE TABLE combine AS SELECT a.subjid, a.trt, a.visid, a.result-b.result AS
  change, b.result AS baseline
  FROM sample(WHERE=(visid>0)) AS a, sample(WHERE=(visid=0)) AS b
  WHERE a.subjid=b.subjid
  ORDER BY subjid, visid;
QUIT;
```

Invoking the %mmrm macro with these parameters produces the following output.

```
%mmrm(indata = combine,
      dep_var = change,
      ind_var = %str(baseline trt visid trt*visid baseline*visid),
      cls_var = %str(trt visid),
      ddfm_methd = kr,
      rep_var = visid,
      subject = subjid,
      cov_list = un#cs#toep,
      lsmn_var = trt*visid,
      sel_methd = order,
      los = 0.05);
```

The EFFECTS, LSMEAN and LSDIFF data sets produced by PROC MIXED code in above macro are stored in the WORK library, and they can be combined to produce the following typical summary table from PROC REPORT:

```
-------------------------------------------------------------------------------------------------------------------
Variable analyzed: A(mmol/L) - Change from Baseline

Tests of Main Effects:
     ----------------------------------------------------------------------------------------------------------------
     Effect                          F-Value          DF        p-value
     -------------------------------------------------------------------
     Baseline                         14.23         1,  16        .002
     Treatment                         0.49         1,  17        .492
     Visit                             0.45         2,  13        .649
     Treatment*Visit                   0.32         2,  14        .732
     Baseline*Visit                    0.27         2,  13        .767
     -------------------------------------------------------------------------------
```

```
-------------------------------------------------------------------------------------------------------------------
                                                  Within Group   LSMean                                Between Group
Visit    Treatment     N     LSMean     SE        p-value*a      Difference     SE        95% CI        p-value*b
-------------------------------------------------------------------------------------------------------------------
1.00                   10    1.90       1.84       .318
                       10    3.28       1.84       .093          -1.38         2.69      (-7.05,4.29)    .613
2.00                   10   -0.17       1.60       .916
                       10    2.24       1.60       .179          -2.41         2.33      (-7.34,2.51)    .316
3.00                    8    2.11       1.67       .228
                        7    1.90       1.81       .315           0.22         2.55      (-5.31,5.74)    .934
-------------------------------------------------------------------------------------------------------------------
N = number of randomized patients with a baseline and post-baseline result at visit;
*a - Within group p-values are from Paired t-tests.
*b - Treatment Comparison: Between group p-value is from 2-Sample t-tests.
MMRM Model: Change from baseline = Baseline + Treatment + Visit + Treatment* Visit + Baseline* Visit.
```

CONCLUSION

This paper on MMRM analysis mainly concludes:
- ➢ It explains how programmers are to be meticulous to convert statistician's intensions into SAS code for selecting best covariance structure for the given model.
- ➢ In Clinical Research Organizations, there is core demand of standardizing this kind of complex task into macro to save time and increase productivity.
- ➢ If user can understand this macro thoroughly as such this can be used repeatedly with the minimum efforts.

The %MMRM macro provided with this paper has been tested under a wide variety of situations, and can be used to streamline work whenever such analysis is required.

REFERENCES

SAS Institute Inc. 2011. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Linga Reddy Baddam
inVentiv Health Clinical
9th Floor, B-Block, Laxmi Cyber City, Kondapur-500084
Hyderabad, Telangana, India.
+91-9652303033
linga.reddy@inventivhealth.com or lingareddy.baddam@gmail.com
http://www.inVentivHealthclinical.com

Sudarshan Reddy Shabadu
inVentiv Health Clinical
9th Floor, B-Block, Laxmi Cyber City, Kondapur-500084
Hyderabad, Telangana, India.
+91-9703137367
sudarshan.reddy@inventivhealth.com or sudarshansas@gmail.com
http://www.inVentivHealthclinical.com

Chunxue Shi
inVentiv Health Clinical
Maryland, USA
410-245-6111
chunxue.shi@inventivhealth.com
http://www.inVentivHealthclinical.com