



Performance Traps in Multiple Imputations Analyses using Mixed Model Repeated Measures

Piotr Karasiewicz, Biometrics, Late-stage
Development, Respiratory and Immunology
(R&I), BioPharmaceuticals R&D,
AstraZeneca, Warsaw, Poland

Problem Statement

- Execution of the following Mixed Model for Repeated Measures (MMRM) code took very long time (weeks!!!) for one of endpoints. It was our starting model - unstructured covariance matrix (type=UN) and Kenward-Roger degrees of freedom approximation (ddfm = kr).
- We executed this analysis using dataset consisting of 100 imputations (PROC MI based) so CPU time for single data imputation scenario had to be multiplied by factor of 100!

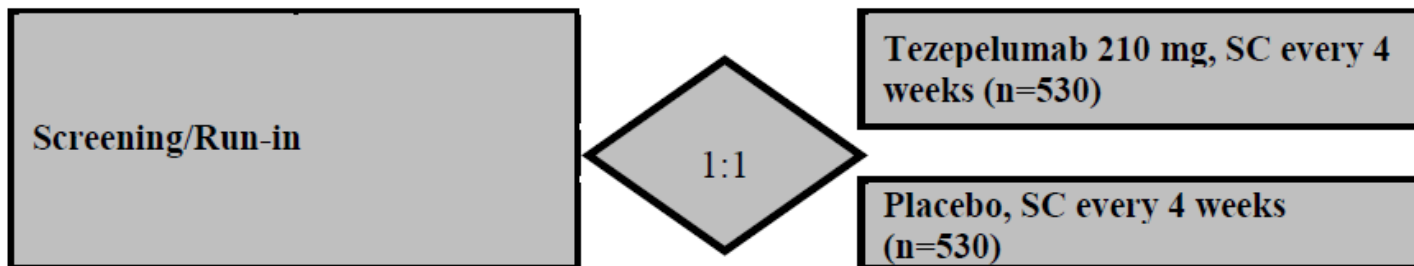
```
proc mixed data = imputed_data method = reml noclprint;  
  class trt01pn avisitn region1n agegr1n;  
  model chg = trt01pn avisitn trt01pn*avisitn region1n agegr1n base /  
          solution ddfm = kr;  
  repeated avisitn / subject = usubjid_num type = UN;  
  lsmeans trt01pn*avisitn / cl pdiff alpha = 0.05 obsmargins;  
  by _imputation_;  
run;
```



Tezepelumab Navigator: study where we faced this challenge

- A Multicenter, Randomized, Double-Blind, Placebo Controlled, Parallel Group, Phase 3 Study to Evaluate the Efficacy and Safety of Tezepelumab in Adults and Adolescents with Severe Uncontrolled Asthma
- Protocol design:

V1	V2-V2a	V3	V4-V16	V17	V18, V19
Day -42 to -35	Day -28 to -25	Week 0	Week 2 to 48	Week 52	Week 58, 64
Screening	Run-in	Randomization	Treatment Phase	End of Treatment	Follow-up



Tezepelumab Navigator: key endpoints

- Primary endpoint: Annualised asthma exacerbation rate (AAER)
- Key Secondary Endpoints:
 - Change from baseline in predose/pre-bronchodilator (pre-BD) forced expiratory volume in 1 second (FEV1)
 - Change from baseline in Standardized Asthma Quality of Life Questionnaire for 12 years and older (AQLQ(S)+12) total score
 - Change from baseline in Asthma Control Questionnaire-6 (ACQ-6) score
 - Change from baseline in weekly mean daily Asthma Symptom Diary score



Asthma Symptom Diary Score (ASD)

- Captured through ePRO device and collected daily from Run-in up to (and including) Week 52
- Questionnaire (0=no Asthma symptoms)

Item	Day-time	Night-time
Severity of wheezing	(0-3)	
Shortness of breath	(0-3)	
Severity of cough	(0-3)	
Severity of chest tightness	(0-3)	
Frequency of waking	N A	(0-3)
Limit activities	(0-3)	NA

- The total daily asthma symptom score was calculated by taking the sum of the daytime score recorded in the evening and the night-time score recorded the following morning and then **7-day (weekly) mean score was derived**.



Selected statistical modelling assumption for key secondary endpoints

Estimation of degrees of freedom for tests of fixed effects: **Kenward-Roger approximation**

Alternative structures of covariance matrix of repeated effects within subject were approached sequentially if convergence criteria is not met for **Unstructured**:

Heterogeneous Toeplitz →

Heterogeneous First Order Autoregressive →

Toeplitz →

First Order Autoregressive →

Compound Symmetry



Statistical modelling – extracts from source documents for key secondary endpoints – **Statistical Analysis Plan (SAP)**

*[...] Unstructured covariance will be assumed to model the relationship between pairs of response variables taken at different visits on the same subject. To allow for the possibility that the MMRM model fails to converge with unstructured covariance, the following hierarchical approach is proposed to select a simpler covariance structure if the preceding covariance structure fails to converge: Unstructured → Heterogeneous Toeplitz → Heterogeneous First Order Autoregressive → Toeplitz → First Order Autoregressive → Compound Symmetry. [...] The **Kenward-Roger approximation** to estimating the degrees of freedom will be used for tests of fixed effects derived from the MMRM model. [...]*



Statistical modelling – extracts from source documents for key secondary endpoints – **Clinical Study Report (CSR)**

[...] This hierarchy was followed for the key secondary endpoints; however, in order to ensure model consistency of the questionnaires for AQLQ(S)+12, ACQ-6, and ASD, the first covariance structure within the hierarchy, which allowed the models for all individual questions, domains and total scores to converge was selected. An exception to this was made to the ASD analyses, where the simpler First Order Autoregressive model was selected directly after the unstructured and heterogeneous Toeplitz models did not converge. The decision was made due to the large amount of diary data, where the analyses were conducted on weekly means for total, daytime, and night-time scores as well as for 6 individual questions. [...]



Why execution of initial MMARM model for ASD took so long?

Covariance matrix aspect (type=UN)

```
proc mixed data = imputed_data method = reml noclprint;  
  class [class variables];  
  model chg = [fixed effects] / solution ddfm = kr;  
  repeated avisitn / subject = usubjid_num type = UN;  
  [...]  
run;
```

- Syntax above implies: no correlation between subjects (as there is no `random` statement) and `repeated` statement means that within each subject we expect correlation between visits. What is more we are not assuming any patterns for this correlation (`type = UN` → unstructured covariance)
- As in our study we had 1000 patients who received treatment and ASD was collected at 52 timepoints the total number of parameters to calculate for correlation matrixes was: $1000 * [(52*53)/2] = 1\ 378\ 000$!!!!



Why execution of initial MMRM model for ASD took so long?

Covariance matrix aspect (type=UN) - continued

- # of parameters for other structures of covariance matrix considered in the study

Structure Name /SAS syntax (Type=...)	# of parameter equation for single Subject	# of parameters to estimate for all Subjects	% of Type=UN
Unstructured / UN	$t*(t+1)/2$	1 378 000	
Heterogeneous Toeplitz / TOEPH	$2*t - 1$	103000	7.47%
Heterogeneous First Order Autoregressive / ARH(1)	$t+1$	53000	3.85%
Toeplitz / TOEP	t	52000	3.77%
First Order Autoregressive / AR(1)	2	2000	0.15%
Compound Symmetry / CS	2	2000	0.15%



Why execution of initial MMARM model for ASD took so long?

Kenward-Roger approximation aspect

```
proc mixed data = imputed_data method = reml noclprint;  
  [...]  
  model chg = [fixed effects] / solution ddfm = kr;  
  repeated avisitn / subject = usubjid_num type = UN;  
  [...]  
run;
```

- Kenward-Roger approximation scales estimator of t or F statistic for fixed effects to reduce level of bias
- This method is recommended for small data samples and adds little value to large samples
- When applied it refits the model multiple times so it increases execution time further
- In combination with unstructure matrix (TYPE=UN) it can inflate execution time significantly



What did we do/consider to improve performance?

- #1 Parallelization of the following `by` statement. 100 analyses were run on separate virtual machines (additional support from IT team was needed) and subsequently we combined analysis results together

```
proc mixed data = imputed_data method = reml noclprint;  
  [...]   
  by _imputation_;  
run;
```

- #2 We hoped we could use multiple CPUs (threads) but **proc mixed** doesn't support this feature.
- #3 We used combination of **proc hpmixed** (some internal algorithms of the procedure improved but still no support for multiple threads) to get the covariance matrix quicker and then **proc plm** to obtain LSMEANS.
- #4 Estimation of initial structure of covariance matrix (for ~10% of patients only) and passing it through PARMs statement (with /NOITER) to final call of **proc mixed** (with more control over convergence criteria via CONVG=absolute option)
- #5 We considered using **proc lmixed** available in SAS Viya. It was very promising as it supports multiple threads but it doesn't support DDFM=KR so turned out not a real option for us



What did we do/consider to improve performance?

SAS code example

- #3 We used combination of **proc hpmixed** (some internal algorithms improved but still no support for multiple threads) to get covariance matrix quicker and then **proc plm** to obtain LSMEANs.
- No visible difference in execution times (executed for one imputation scenario and 24 weeks)

```
proc hpmixed data = imputed_data method = reml noclprint;  
  class trt01pn avisitn region1n agegr1n;  
  model chg = trt01pn avisitn trt01pn*avisitn region1n agegr1n base /  
          solution ddfm = kr;  
  repeated visit / subject = usubjid_num type = UN;  
  store work.output_model  
run;  
  
proc plm restore=work.output_model;  
  lsmeans trt01pn*avisitn / cl pdiff alpha = 0.05 obsmargins=imputed_data;  
  ods output LsMeans=work.lsmeans Diffs=work.diffs;  
run;
```



Going beyond limitations of Statistical Analysis Plan (SAP)

Comparison of execution times for up to week 24 (single imputation) for **1000 patients**

TYPE Option	DDFM Option	Convergence criteria met?	PROC MIXED execution time
UN	KR	Yes	6234.24
TOEPH	KR	Yes	325.5
ARH(1)	KR	Yes	114.1
TOEP	KR	Yes	73.5
AR(1)	KR	Yes	2.1
CS	KR	No	.
UN	residual	Yes	30.3
TOEPH	residual	Yes	152.0
ARH(1)	residual	Yes	58.9
TOEP	residual	Yes	30.4
AR(1)	residual	Yes	1.3
CS	residual	No	.



Exploring DDFM=**residual**. Comparison of covariances

TYPE=UN, DDFM=KR vs. DDFM, up to week 24, 1000 patients

- Estimates (LS Means) are identical
- Lower and Upper confidence intervals differ by 0.0001 which is 0.2% of mean estimate (-0.4851)



Exploring DDFM=residual. Comparison of covariances

TYPE=UN, DDFM=KR vs. DDFM=Residual, up to week 24, 1000 patients

- In `repeated` statement of `proc mixed` RC option was used to generate root of blocks of actual covariance matrix. Based on Cholesky's decomposition matrix which is positive-definite can be presented as product of its lower triangular matrices.
- Those matrices was used for comparion.
- Syntax for `proc calis` :

```
proc calis covpattern=eqcovmat;  
  var COL1 [...] COL24;  
  group 1 / data=cov_kr nobs=24 label="DDFM=KR";  
  group 2 / data=cov_residual nobs=24 label="DDFM=Residual";  
  fitindex NoIndexType On(only)=[chisq df probchi];  
run;
```

- Note: both datasets defined in `data` needs to have type=COV (`data cov_residual (type=cov) ;`)
- Chi-square fit statistic supported not rejecting H0 about equality of matrixes. Similar conclusion was reached when `proc compare` was used.



Conclusions

- When planning MMRM analyses number of parameters in covariance matrix of repeated effects should be taken into consideration to identify issue of “massive models”.
- For massive models Kenward-Roger degrees of freedom approximation paired with unstructured covariance matrix of fixed effects will not be feasible. Even if powerful CPUs are deployed by Sponsor and results are eventually obtained it is very unlikely that Authorities will be able to deploy such demanding hardware. So, we need to think broader!
- DDFM = Residual seems to be reliable method for “big” samples. Additional work needs to be conducted to define a threshold dividing studies into small and big samples.
- Other alternative is using ANCOVA at single time point (e.g. at the last treatment week) as alternative for MMRM.
- PROC CALIS is very handy way method for checking equality of covariance matrices.

