A SIMULATION STUDY USING A MIXED MODEL FRAMEWORK TO ANALYZE THE IMPACT OF SAMPLE SIZE AND VARIABILITY ON TYPE I ERROR

ABSTRACT

Aims: . This simulation study was conducted to check the validity of a MIXED model's statistical inference when violating the underlying assumptions – normality of random errors when there are unbalanced group sizes and inequality of variance of errors [Scheffe, 1959].
Study design: Monte Carlo Simulation Study
Place and Duration of Study: North Dakota State University 2020-2021
Methodology: Repeated measures designs (or longitudinal studies) are commonly seen in many research fields, especially in pharmaceutical clinical trials, agricultural research, and psychology. PROC MIXED (SAS Inc.) is a well-known standard tool for analyzing repeated measures data nowadays. The MIXED procedure is based on the standard linear MIXED model, which estimates parameters by maximizing the restricted likelihood. The usual assumption for a standard linear MIXED model is normality. However, the character of data in the real world may be non- smoothed, or non-symmetric, or having heavy tails. We estimate the Type I error rates in different combinations of settings and compare them with the stated Type I error.
Conclusion: The main results in this study show us that the MIXED model is reasonably robust to modest violations of the normal distribution. However, when a small sample size associated with a treatment was combined with the effects of that treatment having a large variance, a severe inflation problem on Type I error rates could occur when using the MIXED model procedure. When the Type I errors were found to be inflated, the Group= option was found to oftenhelpwiththisproblem. ASub-Samplingprocedure was also found to help with this problem.

Keywords: Repeated Measures; Covariance Structures; Behrens-Fischer problem; Unbalanced Group Sizes; Non-normal distributions*)*

.

1. INTRODUCTION

Repeatedmeasuresdesign(longitudinalstudy)isastudyinwhichtheoutcomevariablesare measuredmorethanonceovertimeforeachsubject.Itiswidelyusedinmanyresearch fields,especiallyinpharmaceuticalclinicaltrials,agriculturalresearch,andpsychology.Thereare three traditional ways to analyze repeated measures data: ANOVA, MANOVA, and MIXED models, notably using SAS PROC MIXED [1]. Among the three, PROC MIXED allows us to specify the variance/covariance structure and tolerate the missing outcome values,makingitastandardtoolforrepeatedmeasuresdatanowadays.

In a repeated measures study, unbalanced samplesize features, unequal group variance features, and non-normal distributions are very common. For example, subjects may dropout during a longitudinal study, which may cause unbalanced group sizes; treatments are likely to have heterogeneous variances; and there are times when the distributions may be skewed instead of normal. In this study, we wish to check the validity of the statistical inference of the MIXED model approach when assumptions are violated [2], We will investigate what happens when we have unbalanced group sizes, non-normal distributions, and inequality of variance of errors in a repeated measures design.

The linear model can be written in matrix form [3]

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}w+\boldsymbol{e} \quad (1.1)$$

When we have T time periods and N subjects, the dependent variable **Y**is a $TN\times1$ vector, representing T time measurements for N subjects. **X** is a TN×2 design matrix containing the average intercept 1 and the slope TIME. $\boldsymbol{\beta}$is a 2×1 vector having two fixed but unknown parameters β_0 and β_1. **Z** is a TN×2N design matrix, and $\boldsymbol{w}$is a 2N×1 vector having the random effects of $w_{0i}$and w_{1i}representing individualsubject'sdifferenceandfollowsN(**0**,**G**).$\boldsymbol{e}$isaTN×1vectorcontainingtherandomeffects formeasurementdifferenceandfollowsN(0,**R**).The$\boldsymbol{w}$and$\boldsymbol{e}$areassumedtobeuncorrelated.

There are two components in equation 1.1: fixed effects $\mathbf{X}\boldsymbol{\beta}$and random effects $\mathbf{Z}\boldsymbol{u}+\boldsymbol{e}$. $\mathbf{X}\boldsymbol{\beta}$is the mean of Y. It is fixed effects because **X** is the design matrix, and the parameter $\boldsymbol{\beta}$can be fixed. There are two kinds of random effects: between-subject random effects $\mathbf{Z}\boldsymbol{w}$, and within-subject random effects $\boldsymbol{e}$. The random-effects $w_{0i}$and w_{1i}in $\boldsymbol{w}$are between-subject variation, representing the deviation of i_{th}subject's intercept and slope from the average intercept and slope. The variable $\boldsymbol{e}$is within-subject random error, where the element $e_{ti}$is the deviation of i_{th}subject at thet_{th} measurement from the subject's individual regression line. The random effects $\boldsymbol{w}$has a covariance matrix **G**, and error $\boldsymbol{e}$ has a covariance matrix **R**.

Since the model contains two random effects, the properties of Y can be investigated by conditioningonrandomeffects.Therefore,thegeneralizedlinearmixedmodelcontainstwotypes of distributions; a conditional distribution given by Equation 1.2; and a marginal distribution given by Equation 1.3, depending on if conditioning on random effects $\boldsymbol{w}$[3]. If there are no random effects($\boldsymbol{w}$=**0**)inthemodel,themarginalandconditionalvariancesareidentical.

The conditional distribution of y with the following mean and variance

$$y|w\sim MVN(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{w},\mathbf{R}) \quad (1.2)$$

Where **R** is a block diagonal matrix with N blocks (one per subject), having dimensions T×T.

The marginal distribution of y with the following mean and variance

$$y\sim MVN(\mathbf{X}\boldsymbol{\beta},\mathbf{V}) \quad (1.3)$$

Where the variance $\boldsymbol{V}$equals to [4]:

$$\mathbf{V}= Var(\mathbf{Z}\boldsymbol{w}+\boldsymbol{e})$$
$$= Var(\boldsymbol{w}) + Var(\boldsymbol{e})$$
$$= \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathbf{T}}+\mathbf{R}$$

Since the data are repeatedly measured, the errors in the mixed model are correlated. A common correlation among measurements is assumed for each subject, so there are multiple choices of covariance structures that can be chosen as a common correlation [1,5,3,6] . We will mention the five most common covariance structures for a repeated measures model.

Variance Components(VC) [3,6] is the default covariance structure for the PROC MIXED procedure in SAS and is also the simplest covariance structure. It has different subject variances in the diagonal and has zero in all off-diagonals. This structure assumes independence oferrors.

First-Order Autoregressive(AR) [3,6] is used widely in time series data. It assumes time intervals between any two repeated measurements are equal. The correlation between two measurements is defined by the exponential function ρ^{x},sothecorrelationwilldecreasewhentime-spaceincreases.

Toeplitz(TOEP)[3,6]hasmoreparametersthanVC,AR(1), and Compound Symmetry (CS),butasmallernumberofparametersthan Unstructured (UN).Themeasurementstakenatclosertime intervals have similar correlations.

Compound Symmetry(CS) [3,5] is used for repeated measures having the same correlation. A constant correlation is assumed between two separate measurements.

Unstructured(UN) [3,5] isthemostcomplexcovariancestructure because each term can be different. It may be the best structure when fitting the real data since the correlation between any two measurements does not have any constraints. However, it mayuseupmanydegreesoffreedomwhichwouldcausetheTypeIerror to increase,especially when the data set is small.

There are lots of guidelines and published papers about how to use SAS PROC MIXED. PROC MIXED is based on REML (restricted maximum-likelihood) approach for parameter estimation [3,7], The F tests are the default statistical tests in PROC MIXED procedure for the main effects, and interaction effects of repeated measures data, which tends to cause Type I error inflation problems with multiple covariance structuresinunbalanceddesigns,andnon-normaldatadistribution[8]. However, the Satterthwaite F test, which can adjust the denominator degrees of freedom of F test through PROC MIXED is fairly robust compared with the default F test on the same condition [9]. Therefore, the DDFM= option in the MODEL statement is important becauseitcanspecifythemethodforcomputingthedenominatordegreesoffreedomforthefixed effects tests. There are five methods for DDFM=, which are CONTAIN, BETWITHIN, RESIDUAL, SATTERTH, and KENWARDROGER. Among the five, DDFM=KENWARDROGERadjuststhedenominatordegreesoffreedombasedonSatterthwaite-typedenominatordegreesoffreedom[10, 11, 12],whichmakesit effectively control the Type I error rate for the repeated measuresfixed-effect.

Based on repeated measurement data, there are two statements in PROC MIXED that needtobespecified:REPEATEDstatementandRANDOMstatement [3].The REPEATED statement can specify the variable name of a repeated measure factor. Within a REPEATEDstatement,theSUBJECToptiondefinesthesetsofrepeatedmeasures,andtheTYPE optionnamesthecovariancestructure,whichmustbeusedwhenonlyusingREPEATEDstatement. The RANDOM statement can specify the random effects. When repeated measures are modeled with a REPEATED statement without a RANDOM statement in PROC MIXED, this model is called a conditional model based on the conditional distribution of Y. In a conditional model, the TYPE option under the REPEATED statement incorporates the complex covariance structure directly through the variance

matrix **R**. When repeated measures are modeled with both REPEATED statementandRANDOMstatementinPROCMIXED,themodeliscalledamarginalmodelbasedonthemarginaldistributionofY.Inthemarginalmodel,theTYPE=optionundertheREPEATED statementspecifiesthevariancematrix**R**whichistypicallydenotedforvariancematrixofrandom error e, and the TYPE= option under the RANDOM statement specifies the variance matrix **G** whichistypicallydenotedforvariancematrixofrandomerror***u***[3].Basedonour simulation experiment results, these two models would provide the same parameter estimates for fixedeffects.

When fitting a model with heterogeneous variance structure, a model with unequal variances can be specified in PROC MIXED under the REPEATED/RANDOM statement with the GROUP= option [3]. The GROUP= option allows the parameters of different GROUPeffectlevelstohavedifferentstructureparametersdespiteacovariancestructure(TYPE= option) remaining the same. It will change the covariance parameters from one group to another,whichcansubstantiallyincreasethenumberofcovarianceparametersneedingtobeestimated [6]. Also, GROUP= option is limited to categorical factors, which requires using a CLASS statement. For example, when incorporating between-subject variance heterogeneity,the GROUP= option in the REPEATED statement can be set up. An example code can be viewed as below [5].

```
proc mixed;
  class A;
model y = A / ddfm=satterth;
repeated /group=A;
lsmeans A / adjust=smm adjdfe=row;
  run ;
```

1.1 Previous Simulation Studies

The Behrens–Fisher problem [13,14,15] has existed for more than sixty years in the area of statistics. The problem is named after Walter Behrens and Ronald Fisher. It occurs when testing the means of two independent populations without knowing the equality of the variances [13,14,15] The Behrens-Fisher problem considers the basic design features, unequal or unknown variances, under two normally distributed populations. However, data in the real world are more often skewed, non-smoothed, non-symmetric, and having heavy tails [16]. The test statistics are not always applied to an ideal situation, like equal sample sizes, equal variances. Therefore, there are lots of studies about the analog of the Behrens-Fisher Problem.

Henry Scheffe talked about the effects of departures from the underlying assumptions in his book “The Analysis of Variance” [2], which is one of the analogous problems of the Behrens-Fisher Problem for the non-normal distribution. In this book, he violates the following assumptions [2]

1. normality of errors, and normality of the random effects in the models;
2. equality of variance of the errors;
3. statistical independence of the errors.

Based on his real data examples, he came up with three conclusions [2]

1. nonnormality has minimal impact on inferences about means but substantial impact on inferences about the variances of random effects;
2. Unequal variance has little impact on inferences about means when sample sizes are equal but has notable impact when sample sizes are unequal;
3. Correlated observations can cause severe problems with inferences about means.

According to the underlying violations mentioned above, some methods are recommended for addressing the severe effects when having two population groups [2,15] With assumed equal size, the classical Student's T-test is recommended. If two populations have equal group size or if the distributions are symmetrical, the Student's t-test is robust. If two populations have unequal group size and the distributions are skewed, the effects of departure from normality may be a concern. If population distributions are normal but with unequal and unknown variances, either Satterthwaite's t-statistic or Satterthwaite's F test is suggested. However, Satterthwaite's procedure is not robust under most non-normal distributions [15,17].

Researchers in another simulation study addressed the robustness of the violations of distribution assumptions and missing values on the estimated of coefficients in linear models [18]. The main concern of this present study is how the Type 1 errors will be affected when conducting hypothesis testing.

1.2 DataTransformation

For equal spreads and reducing skewness of distributions, data transformation is usually the first step to deal with data. A transformation is done to replace a variable with a function of that variable. After a data transformation, the shape of the distribution or relationship will be changed. There are many functions used for data transformation, such as log(x), square (x^2), square root ($x^{0.5}$). Among them, the rank test [18] is one of the standard tools in an applied statistician's tool kit because of its convenience and simplicity. It replaces the original observations with their respective rank, then computes tests on these ranks.

Aligned rank transformation [20] adds a simple alignment fix-up methodology before ranking. The purpose of alignment is to remove the effect of "nuisance" parameters when testing the effects of parameters of interests for multi-parameter models. For example, the effect of blocks in testing for effects of treatments can be removed by data alignment in completely randomized block design [19].

1.3 Sub-SamplingandBootstrapMethod

Sub-sampling and Bootstrap are widespread re-sampling methods. Comparing traditional methods, they require fewer assumptions and are more accurate in practice [21]. Generally speaking, sub-sampling is the method to draw a subset randomly and without replacement from the original data samples [22,23,24] Bootstrap generates samples with replacement randomly from original data samples, usually of the same size as the original sample [25].

2. MATERIAL AND METHODS

2.1 SimulationProgram

To explore the results of statistics mentioned previously, we used SAS 9.4 [5] to perform all simulations and analyses [26] Each simulation was examined using 5000 samples with a 0.05 significance level. There are two intervals used as the index for the estimates' precision: Bradley's liberal criterion [27] and binomial standard error interval [27,28] The test robustness can be evaluated by whether theempiricalTypeIerrorestimate($\hat{\alpha}$)stayswithintheintervalof$0.5\alpha \le \hat{\alpha} \le 1.5\alpha$,whichis $0.025 \le \hat{\alpha} \le 0.075$ inthisstudy.Thebinomialstandarderror is $\sqrt{\left(\frac{\hat{\alpha}(1-\hat{\alpha})}{N}\right)}$ iswhereNisthetotal number of samples. In this study, with a significance level of 0.05 and 5000 samples, the Type I errorrateshouldstaybetween0.04396and0.05604.

2.2 Hypotheses

The Type I error rate was calculated by counting the number of times that the null hypothesis $H_0$was rejected when H_0 is true and dividing by the total number of samples. There are three hypotheses included in this study:

- All Treatment main effect means equal, $H_0 : \tau_1 = \tau_2$
- All Time main effect means equal, $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$
- All Treatment×Time interaction effect meansequal,

  $H_0 : \tau\alpha_{11} = \tau\alpha_{12} = \tau\alpha_{13} = \tau\alpha_{14} = \tau\alpha_{21} = \tau\alpha_{22} = \tau\alpha_{23} = \tau\alpha_{24}$

2.3 DataSimulation

Thissimulationstudywasperformedwith5000samples,andeachsamplewasconducted byasplit-plotdesignassumingequallyspacedtimeintervals.Oursplit-plotdesignhas2treatment groups, 4 repeated time periods, and a First-Order Autoregressive [AR(1)] correlation structure [3,6] where ρ=0.75. There are two stages in this experiment. In the first stage, subjects are randomly assigned to treatment groups (whole-plot factor); In the subsequent stage, time factor (sub-plot factor) in repeated measures is nested within each of the subjects without randomization. There are 30 subjects in each sample. Each subject was randomly assigned to a treatment group and was repeatedly measured four times. For better understanding, one sample data set is visualized in Chart1. The number of subjects in Control or Treatment changes depending on the specific simulationscenarios but initially was split equally with 15 subjects in each group.

Chart 1. Repeated measures data with 30 subjects

Subject ID	Treatment	y1	y2	y3	y4
1	Control	.	.	.	.
2	Control	.	.	.	.
3	Control	.	.	.	.
..	..	..	..	..	..
..	..	..	..	..	..
28	Treatment	.	.	.	.
29	Treatment	.	.	.	.
30	Treatment	.	.	.	.

An effects model for this experiment is

$$Y_{ijk} = \mu_{ij} + \gamma_k + e_{ijk} = \mu + \tau_i + \alpha_j + (\tau\alpha)_{ij} + \gamma_k + e_{ijk} \quad (2.1)$$

Where

Subjects k=1,2, ... ,30 (Subjects 1-15 receive Control and 16-30 receive Treatment)

Treatments i=1,2

Time periods j= 1,2,3,4

$\mu_{ij} = \mu + \tau_i + \alpha_j + (\tau\alpha)_{ij}$is the mean μ for treatment i at time j

τ_i, time effects α_j, and interaction treatment×time effects $(\tau\alpha)_{ij}$, respectively.

$\gamma_k$isthewhole-ploterroreffectforsubjectk,assumed$\overset{id}{\sim}$N(0,σ^2_{γ}).

$e_{ijk}$is the sub-plot error effect for j_{th}time measurement of subject k on treatment i, assuming $\overset{id}{\sim}$N(0,σ^2).

$\gamma_k$and $e_{ijk}$are assumed to be independent of one another

The corresponding matrix form of this model is

$$\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{w}+\boldsymbol{e} \quad (2.2)$$

Where

Y is the vector of observations.

β is the coefficient vector corresponding to the fixed effects μ_{ij}.

X is the design matrix for the fixed effects.

w is the coefficient vector corresponding to whole-plot errors.

Z is the design matrix with respect to whole-plot errors.

To obtain the repeated measures **Y**in a simulation study, two parts of this matrix model needed to be provided. The first part is to specify the fixed effects **X**$\boldsymbol{\beta}$, and the second part is to generate the two random effects - **Z**$\boldsymbol{w}$and $\boldsymbol{e}$, respectively.

The first part, the mean effects u_{ij}in the effects model can be obtained by the fixed unknown constant **X**$\boldsymbol{\beta}$in the matrix model, which contains design matrix **X** and parameter vector $\boldsymbol{\beta}$. Since it isa2by4split-plotdesign,thevector$\boldsymbol{\beta}$issetas$\boldsymbol{\beta}=(\mu,\tau_1,\tau_2,\alpha_1,\alpha_2,\alpha_3,\alpha_4)^t$withnointeraction effects being considered. In this study, the parameter vector $\boldsymbol{\beta}_{7\times1}$was set as $\boldsymbol{\beta}=$ (5, 0, 0, 0, 0, 0, 0)^{t}whentherearenomaineffectsassumed,itwasappliedwhenH_0istrue. The design matrix **X** would have 7 columns that correspond to each parameter in $\boldsymbol{\beta}$, and have 120 rows that correspond to each measurement of each subject. (30)(4)=120 rows because each sample has 30 subject, and 4 repeated measures per subject.

In the second part - two random effects, Ramon Littell [3] provided a formula for getting the random effects variance. That is $\boldsymbol{V}= var(y) = \sigma^2_g\mathbf{J} +\boldsymbol{R}$, where **J**is a matrix of ones. The σ^2_g**J**is the variance for the between-subject random effect **Z**$\boldsymbol{w}$, and the $\boldsymbol{R}$is the variance for the within-subject random effect $\boldsymbol{e}$. The two random effects were both assumed with a mean zero. Therefore,**Z**$\boldsymbol{w}$hasmeanzeroandcovariancematrixσ^2_g**J**,and$\boldsymbol{e}$hasmeanzeroandcovariance matrix **R**.

The part **J**which is a matrix of ones was chosen as between-subject covariance structure because the measures are on the same subject, and $\sigma^2_g$is the variance of treatment groups. The part**R**representsthecovarianceduetotheproximityofmeasurements. **R**isacovariancematrixcorresponding to a within-subject variance. In this study, we assumed that the within-subject variances **R**for all subjects areidentical.

Regardingchoosingacovariancestructurefor**R**,Unstructured (UN) [6] iscommonlyrecommendedastheinitialcovariancestructurewhenusingtheMIXEDmodelforrepeated measures data because the right covariance structure is unknown. However, defining a correlation for any pair of terms would be difficult since there would not need to be any pattern for the Unstructured (UN). Meanwhile, the Unstructured (UN) has the most parameters compared with other structures, which may cause loss of power. Therefore, for obtaining a time series structure, First-Order Auto-regressive (AR(1)) [3,6] was chosen as the right covariance structure with the correlation ρ as 0.75, and the within variance $\sigma^2$was set as 1. The covariance matrix **R**in this study is presented asfollows.

$$R = \sigma 2\begin{pmatrix} 1 & \rho & \rho 2 & \rho 3 \\ \rho & 1 & \rho & \rho 2 \\ \rho 2 & \rho & 1 & \rho \\ \rho 3 & \rho 2 & \rho & 1 \end{pmatrix} = \sigma 2\begin{pmatrix} 1 & 0.75 & 0.5625 & 0.421875 \\ 0.75 & 1 & 0.75 & 0.5625 \\ 0.5625 & 0.75 & 1 & 0.75 \\ 0.421875 & 0.5625 & 0.75 & 1 \end{pmatrix}$$

2.3.1 Simulation Scenarios

The usual assumption for a standard linear MIXED model is normality.For checking the validity of the MIXED model, we violated the assumption by simulating normal/non-normal distribution of between-subject effects ***Zw***and within-subject effects ***e***[2] specifically. Therefore,fourdifferentscenariosweregenerated:

1. betweensubjecteffects***Zw***followsamultivariatenormaldistributionandwithin-subjecteffects***e***
2. followsmultivariatenon-normaldistribution

   Z*w*∼MVN(**0**, σ^2_g**J**), ***e***∼Multi-Skew(μ=**0**, **R**, Skew=2,Kurtosis=6)

3. between-subject effects ***Zw***follows a multivariate normal distribution and within-subject effects ***e***follows multivariate normal distribution

   Z*w*∼MVN(**0**, σ^2_g**J**), ***e***∼MVN(**0**,**R**)

4. between-subject effects *Zw* follows multivariate non-normal distribution and within-subject effects ***e***follows multivariate normal distribution

   Z*w*∼Multi-Skew(μ=**0**, σ^2_g**J**, Skew=2, Kurtosis=6), ***e***∼MVN(**0**,**R**)

5. between-subject effects ***Zw***follows multivariate non-normal distribution and within-subject effects ***e***multivariatenon-normaldistribution

   Z*w*∼Multi-Skew(μ=**0**,σ^2_g**J**,Skew=2,Kurtosis=6),
   e∼Multi-Skew(μ=**0**,**R**,Skew=2,Kurtosis=6)

For simulating the multivariate non-normal distribution data in this study, the univariate distributionwasfirstgeneratedbyFleishman'sCubicTransformation [29] withtargetvalues ofskewness=2andkurtosis=6,thenthemethodfromVale-Maurelli[30]was used to generate multivariate non-normaldata.

Within each scenario, two conditions were applied for checking the stability of the MIXED model. They are the equality of sample sizes, and the equality of (between-subject) variances for two treatment groups[2]. The parameter sets for two conditions were listed as follows. Here, 1 represents the treatment group, and 2 represents the control group.

- Equal group size n_1= n_2=15
- Unequal group size n_1/n_2= (0.5, 2),where
  - For n_1/n_2=0.5, n_1= 10 and n_2=20
  - For n_1/n_2=2.0, n_1= 20 and n_2=10
- Equal variancesσ^2_{g1}= σ^2_{g2}= (1, 2, 4, 10)
- Unequal variances σ^2_{g1} / σ^2_{g2}= (2,4,6,8,10), where σ^2_{g2}=1

The two conditions resulted in four different situation combinations: equal group size with equal variance, equal group size but unequal variance, unequal group size but equal variance, unequal group size and unequal variance.

2.3.2 Analysis

After these data were generated based on these simulation scenarios and different conditions, PROC MIXED was applied to run the test by sample. In PROC MIXED, the conditional distribution of the mixed model was used. DDFM=KENWARDROGER is used to adjust the degrees of freedom. The five most common covariance structures (Variance Components (VC), First-OrderAutoregressive(AR(1)),Toeplitz(TOEP),CompoundSymmetry(CS),Unstructured(UN) [3,6] were applied under the REPEATED statement, respectively. Anexamplecodecanbeviewedasbelow [3]

```
proc mixed  data=rmuv;
```

```
by sample;
class trt period subj_id;
model stress = trt | period / ddfm=kr;
repeated period / subject=subj_id  type=AR(1);
title2 "Repeated Measures ANOVAusing Mixed ModelApproachAR(1)";
run;
```

3. RESULTS AND DISCUSSION

We estimate Type I error rates by finding the percentage of the cases that reject the null hypothesis H_0 when H_0 is true. Also, the significance level α is stated as 5.00%. The Type I error rate for all combinations of two conditions: the equality of group size and the equality of variance, under four basic scenarios are presented in the following tables in percentage form. In eachtable,therearethreetestpartsofdifferenteffects:treatmenteffects,timeperiodeffects,and interaction (treatment×time) effects.

3.1 Four Scenarios

The four distribution scenarios that we are considering are listed as follows:

1. $\boldsymbol{Zw}$~ Normal &$\boldsymbol{e}$~Skew
2. $\boldsymbol{Zw}$~ Normal &$\boldsymbol{e}$~Normal
3. $\boldsymbol{Zw}$~ Skew &$\boldsymbol{e}$~Normal
4. $\boldsymbol{Zw}$~ Skew &$\boldsymbol{e}$~Skew

Based on the parameters we set, the first two scenarios have normal distributions or distributions that are slightly skewed, and the last two scenarios have distributions that are heavily skewed.

Table1displaystheTypeIerrorratesoffourbasicscenariosfortreatmentandcontrol group with $n_1 = n_2 = 15$ and equal variances$\sigma^2_{g1} = \sigma^2_{g2}$that are increasing (1,2,4,10). The Type I error rates in both period results and interaction results are below the limits of α equals 5.00%, staying between 3.36% and 4.88%, a bit below the binomial threshold (4.396, 5.604). Values of the Type I error rates in treatment results stay within 4.44% to 5.48%, but the highest value, 5.48%, is still below the upper bound of binomial standard error interval (4.396, 5.604). Also, the tests are robust here. The four scenarios have a similar trend, so the violation of normality appears to cause little effect on Type I error rates no matter whether the skewness is in between-subject effects or within-subject effects.

In Table 2, results are given for unequal sample sizes but equal variances. In Table 3, results are given for equal sample sizes, but unequal variances. With only unequal variance, the Type I error rates in period, and interaction tests are all below the limit of 5.00%. However, in the treatment test, when increasing the variance ratio, when the distributions are normal or slightly skewed, the Type I error rate stays within 95% binomial standard error interval (4.396, 5.604), but when the distributions are highly skewed the Type I error rates increase from 4.66% to 7.86% Therefore, unbalanced group size or unequal group variance itself should not be a concern when using a MIXED model. However, when the distributions are highly skewed and for unbalanced data,theeffectsofabigdifferencebetweengroupvariancesshouldbeaconcern.

Table 1. Type I error Rate of Four Basic Scenarios for balanced group size $n_1 = n_2 = 15$ and equalvariances$\sigma^2_{g1} = \sigma^2_{g2} = (1, 2, 4, 10)$.

σ_γ^2	Distribution Scenarios			
	Zw~ Normal **e**~ Skew	**Zw**~ Normal **e**~ Normal	**Zw**~Skew **e**~Normal	**Zw**~ Skew **e**~ Skew
Treatment Results				
1	4.96	5.48	5.38	5.12
2	4.80	4.66	5.04	4.78
4	5.34	5.46	5.44	4.90
10	4.96	5.14	4.44	4.52
Period Results				
1	3.74	4.20	4.86	3.76
2	3.50	4.12	4.14	3.64
4	3.70	3.84	4.34	4.10
10	3.98	4.58	3.96	3.66
Treatment*Period Results				
1	3.52	3.54	4.22	3.70
2	3.36	4.54	4.02	3.90
4	3.42	4.28	3.86	3.84
10	3.84	4.88	4.02	3.66

Table2.TypeIerrorRateofFourScenariosforunequalgroupsizen_1=10andn_2=20 and equal variancesσ^2_{g1}= σ^2_{g2}= (1, 2, 4, 10).

σ_γ^2	Distribution Scenarios			
	Zw~ Normal **e**~ Skew	**Zw**~ Normal **e**~ Normal	**Zw**~Skew **e**~Normal	**Zw**~ Skew **e**~ Skew
Treatment Results				
1	5.06	4.72	5.14	5.16
2	5.34	5.96	4.98	4.24
4	5.72	4.70	4.48	4.88
10	4.78	5.10	4.38	4.40
Period Results				
1	4.14	4.66	4.16	4.26
2	4.18	3.94	4.52	3.82
4	3.72	4.14	4.30	3.96
10	4.18	4.04	4.62	3.72
Treatment*Period Results				
1	4.04	3.64	4.34	3.94
2	4.04	4.46	4.24	3.84
4	4.34	4.24	4.12	3.84
10	3.96	3.62	4.00	4.26

Table 3. Type I error Rate of Four Scenarios for equal group sizes n_1 = n_2 = 15 and unequal variance as giving the variance ratio $\sigma^2_{g1}/\sigma^2_{g2}$=(2,4,6,8,10), where σ^2_{g2} =1.

	Distribution Scenarios

$\sigma^2_{g1}/\sigma^2_{g2}$	**Zw**~ Normal **e**~ Skew	**Zw**~ Normal **e**~ Normal	**Zw**~Skew **e**~Normal	**Zw**~ Skew **e**~ Skew
Treatment Results				
2	5.32	5.36	5.44	4.66
4	5.16	5.42	5.34	6.28
6	5.34	5.20	6.58	6.32
8	5.08	5.28	7.28	6.90
10	5.24	5.04	7.86	7.76
Period Results				
2	4.02	4.66	3.94	3.96
4	3.58	4.00	4.12	4.00
6	4.04	3.94	3.66	3.20
8	3.46	4.28	3.90	4.02
10	3.96	4.20	3.82	3.88
Treatment*Period Results				
2	3.84	4.00	3.96	3.94
4	3.72	4.14	4.24	3.80
6	3.70	4.12	4.40	3.66
8	3.66	4.34	3.66	3.52
10	3.44	4.04	4.30	3.56

Table 4 illustrates how the Type I error rate varies when we have unequal sample sizes and unequal variances at the same time. The ratio of the two sample sizes equal to 1 is provided as a reference. In period and interaction tests, Type I error rates stays below the limit of 5.00%. Under the treatment test, when the size ratio equals to 2, Type I error rates are conservative as the variance ratio increases in all four scenarios, namely they stay below 5.00. When the size ratio equals to 0.5, the Type I error rates are inflated incredibly as the variance ratio increases in all four scenarios, rising from 6.36% to 15.02%. This is above the upper bound of 95% binomial interval (4.396, 5.604). It also means that most treatment tests are not robust when the size ratio equals to 0.5. Meanwhile, all four distribution scenarios have a similar pattern in all size ratios except for size ratio one. According to these results, when small group size combines with large variance, it would cause a severe inflation problem on Type I error rates, which breaks the MIXED model'sperformance.

Therefore, we came up with two conclusions in this part:

1. the MIXED model is reasonably robust to modest violations of the normal distribution.
2. when a large variance ratio (greater than 8) combines with heavily skewed distributions with equal sample sizes, the MIXED model cannot be considered robust anymore. Nevertheless, it should not be a concern since the real data usually would not have such a big variance ratio.
3. when there is a small sample combined with large variance, it will cause serious TypeIerrorinflationproblemsthatneedtobepaidattentionto.

Table 4. Type I error Rate of Four Scenarios for different size ratio n_1/n_2 = (0.5, 1, 2), where the total group size is 30; unequal variance as giving variance ratio $\sigma^2_{g1}/\sigma^2_{g2}$ =(2,4,6,8,10), where σ^2_{g2} =1.

	Distribution Scenarios

$\sigma^2_{g1}/\sigma^2_{g2}$	**Zw**~ Normal **e**~ Skew			**Zw**~ Normal **e**~ Normal			**Zw**~Skew **e**~Normal			**Zw**~Skew **e**~skew		
	n_1/n_2			n_1/n_2			n_1/n_2			n_1/n_2		
	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
Treatment Results												
2	7.14	5.32	3.70	6.82	5.36	3.88	6.36	5.44	3.62	6.94	4.66	3.86
4	10.56	5.16	2.42	10.06	5.42	2.54	10.82	5.34	2.96	10.18	6.28	3.30
6	11.10	5.34	2.04	11.06	5.20	1.86	12.60	6.58	2.58	12.12	6.32	3.06
8	12.70	5.08	1.84	12.34	5.28	1.70	13.36	7.28	3.38	13.40	6.90	3.20
10	14.38	5.24	1.34	14.08	5.04	1.90	15.02	7.86	3.00	14.66	7.76	2.52
Period Results												
2	3.96	4.02	4.46	4.14	4.66	4.32	4.16	3.94	4.46	4.36	3.96	3.94
4	4.10	3.58	3.94	3.70	4.00	4.68	4.26	4.12	4.14	4.30	4.00	3.76
6	4.20	4.04	3.80	4.14	3.94	4.52	3.92	3.66	4.38	4.10	3.20	4.34
8	4.02	3.46	4.42	3.90	4.28	4.08	4.24	3.90	4.22	4.56	4.02	4.34
10	4.56	3.96	3.86	4.08	4.20	4.18	4.26	3.82	4.02	3.98	3.88	3.74
Treatment*Period Results												
2	3.98	3.84	4.50	4.62	4.00	3.84	3.90	3.96	4.18	3.78	3.94	3.78
4	3.94	3.72	3.76	4.16	4.14	3.86	4.38	4.24	3.88	4.00	3.80	3.74
6	4.16	3.70	4.06	4.12	4.12	4.32	4.56	4.40	4.22	3.86	3.66	4.02
8	4.02	3.66	3.78	4.70	4.34	3.94	4.02	3.66	3.98	3.94	3.52	4.38
10	3.68	3.44	3.72	3.74	4.04	4.06	3.74	4.30	4.20	4.24	3.56	3.76

3.2 StabilityofTypeIerrorRates

For checking the stability of the Type I error rates, we would increase the sample sizes and the number of repeated time points to see if the Type I error rates keep the same consistency.

3.2.1 Increasing SampleSizes

Inthissection,wechoseonescenario,***Zw***~Normal***e***~Skew,asanexample,andgenerated three different sample sizes under each size ratio. The result of Type I error rates is presented in Table 5. We can see that the Type I error rates keep the same trend under the same size ratio. For example, when the size ratio equals to 0.5 and the variance ratio increases from 2 to 10, the Type I error rate increases from 6.86% to 14.38% regardless of the specific sample size of n_1 and n_2. Also, the Type I error rates have the similar values when they are in the same size ratio and variance ratio. For example, when size ratio equals to 0.5 and variance ratio equals to 10, the Type I error rate keeps around 13 nomatterthedifferenceofsamplesizes:13.18%whenn_1:n_2=20:40,12.98%whenn_1:n_2=30:60,and 13.54% when n_1:n_2=40:80. It shows us that the Type I error rates across the same sample size ratio wereveryconsistent.

Table 5. Type I error Rates of different sets of sample sizes under **ZW** MVN &***e*** exp scenario; unequal variance ratio $\sigma^2_{g1}/\sigma^2_{g2}$=(2,4,6,8,10), where σ^2_{g2}=1.

$\sigma^2_{g1}/\sigma^2_{g2}$	Distribution Scenario: **ZW**~MVN ***e***~skew		
	n_1/n_2=0.5	n_1/n_2=1	n_1/n_2=2

	10:20	20:40	30:60	40:80	15:15	30:30	45:45	60:60	20:10	40:20	60:30	80:40
Treatment Results												
2	7.14	6.98	6.86	6.96	5.32	5.40	5.10	5.98	3.70	3.96	3.50	4.08
4	10.56	10.38	9.08	9.64	5.16	5.46	5.52	5.48	2.42	2.00	2.08	2.46
6	11.10	11.20	11.04	10.98	5.34	5.44	5.30	5.10	2.04	1.90	1.86	2.08
8	12.70	12.12	13.00	12.06	5.08	5.64	5.02	5.04	1.84	1.48	1.22	1.62
10	14.38	13.18	12.98	13.54	5.24	5.64	5.90	4.68	1.34	1.34	1.38	1.34
Period Results												
2	3.96	4.04	4.82	4.42	4.02	4.50	4.02	4.74	4.46	4.70	4.12	4.28
4	4.10	4.22	4.90	4.18	3.58	4.70	4.26	4.70	3.94	4.38	4.64	4.86
6	4.20	4.72	4.60	4.36	4.04	4.68	4.54	4.82	3.80	4.58	4.14	4.34
8	4.02	4.56	4.56	4.68	3.46	4.00	5.20	4.78	4.42	4.88	4.26	5.12
10	4.56	4.30	4.50	4.72	3.96	4.42	4.54	4.98	3.86	4.10	4.40	4.74
Treatment*Period Results												
2	3.98	4.36	4.58	4.34	3.84	4.06	4.56	4.72	4.50	4.62	4.34	4.76
4	3.94	4.16	4.64	4.86	3.72	4.42	4.00	4.20	3.76	4.58	4.66	4.60
6	4.16	4.36	4.94	4.24	3.70	4.40	3.98	4.68	4.06	4.80	4.58	4.68
8	4.02	4.72	4.76	4.38	3.66	3.84	4.64	4.64	3.78	4.54	4.82	4.40
10	3.68	4.24	4.60	5.16	3.44	3.96	5.02	4.84	3.72	4.90	4.84	4.40

3.2.2 Increasing TimePoints

A real-world longitudinal study is likely to have more than four repeated measures, so the trend consistency of Type I error rates for a different number of times points is also essential. Therefore, in this section, we extended the number of time points from 4 to 6 to check Type I error rates. Based on Table 6, we can see that in the treatment test, Type I error rates inflated from 6.94% to 15.68% when the size ratio equals to 0.5, and deflated below 5.00 when the size ratio equals to 2. The trend is the same as Table 3.4 when repeated time points are 4; And in period and interaction tests, Type I error rates all stay below the limit of 5.00. The difference of Type I Error Rate from 6 time points to 4 time points is presented in Table 3.7. The values in Table 3.10 arearound0,whichclearlyshowsusthatthedifferenceisquitesmall. We believe this also gives us more confidence that we can look at ratios of sample sizes and ratios of variances without too much concern for the number of repeated time points.

Table 6. Type I error Rate of Four Scenarios with 6 time points for different size ratio n_1/n_2 = (0.5, 1, 2), where the total group size is 30;unequal variance ratio $\sigma^2_{g1}/\sigma^2_{g2}$=(2,4,6,8,10), where σ^2_{g2}=1.

$\sigma^2_{g1}/\sigma^2_{g2}$	Distribution Scenarios			
	Zw~ Normal ***e***~ Skew	***Zw***~ Normal ***e***~ Normal	***Zw***~Skew ***e***~Normal	***Zw***~Skew ***e***~skew

	n_1/n_2			n_1/n_2			n_1/n_2			n_1/n_2		
	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
Treatment Results												
2	7.42	5.46	3.92	7.50	4.92	4.16	7.66	5.58	3.52	6.94	5.80	3.96
4	9.92	6.04	2.26	9.32	5.98	2.26	9.66	6.08	3.40	10.40	5.92	3.22
6	11.56	5.34	1.92	12.24	5.20	1.70	13.18	6.78	2.84	13.22	6.84	3.42
8	13.32	5.72	1.56	12.98	6.36	1.58	14.86	7.02	3.14	13.88	7.16	3.54
10	14.12	5.40	1.36	14.64	6.06	1.52	15.68	7.30	3.20	14.38	7.44	3.06
Period Results												
2	3.84	3.70	3.92	3.82	4.24	4.22	3.72	4.46	3.96	3.98	3.96	4.06
4	4.16	3.30	4.10	3.62	4.06	3.98	3.88	4.40	3.80	4.28	3.86	4.02
6	4.60	3.72	4.14	4.04	4.38	4.02	4.32	3.90	4.18	3.70	4.06	4.46
8	3.94	3.74	4.04	4.02	4.36	3.78	3.88	4.38	3.70	4.08	3.48	3.90
10	4.12	3.90	3.96	3.24	3.86	3.86	4.24	4.14	3.98	4.00	3.64	4.20
Treatment*Period Results												
2	3.96	3.68	4.70	4.16	3.96	3.76	4.42	4.08	4.14	4.26	3.78	4.00
4	4.08	3.80	4.10	4.36	4.20	3.68	4.46	4.42	3.84	3.70	3.40	3.76
6	4.34	3.52	4.16	4.10	4.18	3.96	3.58	3.84	4.06	4.32	2.94	3.88
8	3.80	3.70	3.88	3.94	4.28	3.82	4.10	4.36	3.72	4.54	3.34	4.02
10	4.08	3.74	3.82	3.94	4.08	4.04	4.22	4.36	3.68	4.04	3.76	3.86

3.3 Five Different CovarianceStructures

In this paper, we mainly focus on two problems presented in the previous two sections: Section 3.1 and Section 3.2. We also want to provide a general idea of how the incorrect covariance structure affects the Type I error rate.

To recall, First-Order Autoregressive (AR(1)) [3,6] was chosen as the correct covariance structure when we generated these datasets. After these datasets were generated,fivemostcommoncovariancestructure(First-OrderAutoregressive(AR(1)),Toeplitz (TOEP),CompoundSymmetry(CS),Unstructured(UN),VarianceComponents(VC) [3,6] were applied under the REPEATED statement in PROC MIXED to run the test, respectively.Therefore,therearefivedifferentTypeIerrorrates,andpowerratescorresponding to each covariance structure were obtained ineverysituation. Nevertheless, only the resultsofTypeIerrorratesunderFirst-OrderAutoregressive(AR(1))structure [3,6] is the correct one, which would be used as the reference for the results under other covariance structures. To sum up, according to all the tables listed below, the Type I error rates under Toeplitz (TOEP), Compound Symmetry (CS), and Unstructured (UN) are very similar to the results under First-Order Autoregressive (AR(1)) structure: the difference among the four is around 1 percent. However, the Variance Components (VC) structure has the worst results among the five: the Type I error rate is 3 to even 17 times than the Type I error rate under First-Order Autoregressive (AR(1)) structure.

Table 7. The difference of Type I Error Rate from 6 time points to 4 time points.

$\sigma^2_{\gamma 1}/\sigma^2{}_2$	Distribution Scenarios			
	Z*w*~ Normal ***e*~** Skew	**Z*w*~** Normal ***e*~** Normal	**Z*w*~**Skew ***e*~**Normal	**Z*w*~**Skew ***e*~**skew
	n_1/n_2	n_1/n_2	n_1/n_2	n_1/n_2

	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
Treatment Results												
2	0.28	0.14	0.22	0.68	-0.44	0.28	1.30	0.14	-0.10	0.00	1.14	0.10
4	-0.64	0.88	-0.16	-0.74	0.56	-0.28	-1.16	0.74	0.44	0.22	-0.36	-0.08
6	0.46	0.00	-0.12	1.18	0.00	-0.16	0.58	0.20	0.26	1.10	0.52	0.36
8	0.62	0.64	-0.28	0.64	1.08	-0.12	1.50	-0.26	-0.24	0.48	0.26	0.34
10	-0.26	0.16	0.02	0.56	1.02	-0.38	0.66	-0.56	0.20	-0.28	-0.32	0.54
Period Results												
2	-0.12	-0.32	-0.54	-0.32	-0.42	-0.10	-0.44	0.52	-0.50	-0.38	0.00	0.12
4	0.06	-0.28	0.16	-0.08	0.06	-0.70	-0.38	0.28	-0.34	-0.02	-0.14	0.26
6	0.40	-0.32	0.34	-0.10	0.44	-0.50	0.40	0.24	-0.20	-0.40	0.86	0.12
8	-0.08	0.28	-0.38	0.12	0.08	-0.30	-0.36	0.48	-0.52	-0.48	-0.54	-0.44
10	-0.44	-0.06	0.10	-0.84	-0.34	-0.32	-0.02	0.32	-0.04	0.02	-0.24	0.46
Treatment*Period Results												
2	-0.02	-0.16	0.20	-0.46	-0.04	-0.08	0.52	0.12	-0.04	0.48	-0.16	0.22
4	0.14	0.08	0.34	0.20	0.06	-0.18	0.08	0.18	-0.04	-0.30	-0.40	0.02
6	0.18	-0.18	0.10	-0.02	0.06	-0.36	-0.98	-0.56	-0.16	0.46	-0.72	-0.14
8	-0.22	0.04	0.10	-0.76	-0.06	-0.12	0.08	0.70	-0.26	0.60	-0.18	-0.36
10	0.40	0.30	0.10	0.20	0.04	-0.02	0.48	0.06	-0.52	-0.20	0.20	0.10

The main tables in the previous sections are presented under 5 different covariance structures in this section, and the guidance for the comparison is listed as below. Table 8 is the TypeIerrorratesunder5differentcovariancestructuresoffourscenariosforthreedifferentsize ratios (n1/n2 = (0.5, 1, 2)) and five different variance ratios ($\sigma^2_{\gamma 1}/\sigma^2{}_2$= (2, 4, 6, 8, 10)), which corresponding to the Treatment Test in Table 4. Table 9 compares different test methods (Original, Rank, Aligned Rank, Sub-Sampling, Group=option) on Type I error rates using 5 different covariance structures of four scenarios for the fixed size ratio (n1/n2 = 0.5) and five different variance ratios ($\sigma^2_{\gamma 1}/\sigma^2{}_2$= (2, 4, 6, 8, 10)), which corresponding to Table 4. In this case, the results of the methods Type I error rates are shown for each type of covariance structure (with AR(1) being correct). The results are divided up into results based on the original data, the results using the Rank transformation, the results using the Aligned Rank transformation, the results using the Sub-Sampling Method, and the results using the GROUP= option in the MIXED model. The TYPE I errors under the VC covariance structure were inflated under all scenarios. The Rank transformation and the Aligned Rank transformation did not maintain Type I errors as well as the Sub-Sampling method and the GROUP= option under the MIXED model. Table 10 compares these methods on Type I error rates for three treatment groups in Treatment Test under 5 different covariance structures for the fixed sample size (n_A= 10; n_B= 10; n_C= 20) and the group variance ratios ($\sigma^2_{\gamma A}$ = 2; $\sigma^2{}_B$= (2, 4, 6, 8, 10); $\sigma^2{}_C$= 1). The error rates for the Rank transformation and the Aligned Rank transformation were not considered here since they were higher than the other two methods for two treatments. The methods of Type I error rates for three treatment groups in Treatment Test under 5 different covariance structures for the fixed sample size (n_A= 10; n_B= 10; n_C= 20) and the group variance ratios ($\sigma^2_{\gamma A}$ = 6; $\sigma^2{}_B$= (2, 4, 6, 8, 10); $\sigma^2{}_C$= 1) were also considered, but not given here, since the results were similar to the results in Table 10.

Table 8. Type I error Rate in Treatment Test under 5 Covariance Structures of Four Scenarios for different size ratio n_1/n_2= (0.5, 1, 2), where the total group size is 30; unequal variance as giving variance ratio $\sigma^2_{\gamma 1}/\sigma^2{}_2$=(2,4,6,8,10), where $\sigma^2{}_2$=1.

$\sigma^2_{\gamma 1}/\sigma^2_{\gamma 2}$	Distribution Scenarios											
	Zw~ Normal **e**~ Skew			**Zw**~ Normal **e**~ Normal			**Zw**~Skew **e**~Normal			**Zw**~Skew **e**~skew		
	n_1/n_2			n_1/n_2			n_1/n_2			n_1/n_2		
	0.5	1	2	0.5	1	2	0.5	1	2	0.5	1	2
First-Order Autoregressive (AR(1))												
2	7.14	5.32	3.70	6.82	5.36	3.88	6.36	5.44	3.62	6.94	4.66	3.86
4	10.56	5.16	2.42	10.06	5.42	2.54	10.82	5.34	2.96	10.18	6.28	3.30
6	11.10	5.34	2.04	11.06	5.20	1.86	12.60	6.58	2.58	12.12	6.32	3.06
8	12.70	5.08	1.84	12.34	5.28	1.70	13.36	7.28	3.38	13.40	6.90	3.20
10	14.38	5.24	1.34	14.08	5.04	1.90	15.02	7.86	3.00	14.66	7.76	2.52
Toeplitz (TOEP)												
2	6.72	5.26	3.56	6.52	5.18	3.70	6.16	5.22	3.42	6.60	4.34	3.58
4	10.04	5.00	2.32	9.86	5.24	2.46	10.54	5.14	2.52	9.74	6.02	3.16
6	10.74	5.10	2.04	10.78	4.98	1.78	12.42	6.40	2.84	11.62	6.04	2.84
8	12.38	4.96	1.72	12.14	5.12	1.62	13.00	7.06	3.12	13.00	6.68	2.96
10	13.80	5.16	1.28	13.76	5.00	1.86	14.60	6.82	2.94	14.34	7.58	2.48
Compound Symmetry (CS)												
2	6.86	5.28	3.62	6.50	5.30	3.68	6.10	5.06	3.46	6.80	4.38	3.84
4	10.16	5.10	2.20	9.92	5.32	2.32	10.56	5.10	2.54	9.84	6.04	3.16
6	10.88	5.18	1.96	11.06	4.92	1.78	12.16	6.50	2.84	11.72	5.96	2.84
8	12.38	5.02	1.68	12.04	5.08	1.60	12.80	7.02	3.18	13.12	6.84	3.04
10	13.92	5.18	1.26	13.64	5.04	1.84	14.56	6.86	2.90	14.36	7.68	2.50
Unstructured (UN)												
2	6.86	5.28	3.62	6.50	5.30	3.68	6.10	5.06	3.46	6.80	4.38	3.84
4	10.16	5.10	2.20	9.92	5.32	2.32	10.56	5.10	2.54	9.84	6.04	3.16
6	10.88	5.18	1.96	11.06	4.92	1.78	12.16	6.50	2.84	11.72	5.96	2.84
8	12.38	5.02	1.68	12.04	5.08	1.60	12.80	7.02	3.18	13.12	6.84	3.04
10	13.92	5.18	1.26	13.64	5.04	1.84	14.56	6.86	2.90	14.36	7.68	2.50
Variance Components (VC)												
2	33.64	30.16	26.26	32.36	30.44	26.92	33.16	30.74	28.00	34.06	30.44	27.60
4	39.74	30.82	23.52	37.92	31.28	24.84	40.00	32.52	22.74	39.60	32.74	23.90
6	10.30	31.36	22.22	39.42	30.64	21.52	41.72	32.86	23.58	42.72	31.40	23.00
8	42.66	31.66	20.90	41.84	31.52	22.04	44.88	33.34	23.30	44.20	32.78	21.80
10	45.62	32.26	19.90	44.44	31.90	22.00	45.36	34.08	21.72	45.22	33.26	22.00

Table 9.Methods for Type I error rates in Treatment Test under 5 Covariance Structures of four scenarios with fixed size ratio n_1/n_2=0.5 and $\sigma^2_{\gamma 1}/\sigma^2{}_2$= (2,4,6,8,10), where $\sigma^2{}_2$=1

$\sigma^2_{\gamma 1}$ / $\sigma^2_{\gamma 2}$	Distribution Scenarios																			
	Zw ~ Normal $\boldsymbol{e}$~ Skew					Zw ~ Normal $\boldsymbol{e}$~ Normal					Zw ~Skew $\boldsymbol{e}$~Normal					Zw ~ Skew $\boldsymbol{e}$~ Skew				
	AR(1)	TOEP	CS	UN	VS	AR(1)	TOEP	CS	UN	VS	AR(1)	TOEP	CS	UN	VS	AR(1)	TOEP	CS	UN	VS
Original Test																				
2	7.14	6.72	6.86	6.86	33.64	6.82	6.52	6.50	6.50	32.36	6.36	6.16	6.10	6.10	33.16	6.94	6.60	6.80	6.80	34.06
4	10.56	10.04	10.16	10.16	39.74	10.06	9.86	9.92	9.92	37.92	10.82	10.54	10.56	10.56	40.00	10.18	9.74	9.84	9.84	39.60
6	11.10	10.74	10.88	10.88	10.30	11.06	10.78	11.06	11.06	39.42	12.60	12.42	12.16	12.16	41.72	12.12	11.62	11.72	11.72	42.72
8	12.70	12.38	12.38	12.38	42.66	12.34	12.14	12.04	12.04	41.84	13.36	13.00	12.80	12.80	44.88	13.40	13.00	13.12	13.12	44.20
10	14.38	13.80	13.92	13.92	45.62	14.08	13.76	13.64	13.64	44.44	15.02	14.60	14.56	14.56	45.36	14.66	14.34	14.36	14.36	45.22
Rank Test																				
2	6.52	6.32	6.36	6.36	32.54	6.16	5.82	5.84	5.84	30.64	6.46	5.96	6.10	6.10	29.94	7.62	7.42	7.38	7.38	33.30
4	9.32	9.10	8.94	8.94	37.10	8.32	8.12	8.02	8.02	34.20	10.66	10.26	10.34	10.34	37.40	11.94	11.44	11.36	11.36	39.26
6	9.38	8.96	9.02	9.02	36.38	9.16	8.76	8.64	8.64	35.72	13.50	13.00	13.14	13.14	42.20	15.48	15.06	15.02	15.02	44.68
8	10.12	9.88	9.80	9.80	37.92	9.90	9.48	9.44	9.44	37.50	16.08	15.42	15.54	15.54	45.50	16.74	16.60	16.50	16.50	46.16
10	11.66	11.26	11.28	11.28	41.06	10.30	10.08	10.00	10.00	38.48	18.24	17.74	17.66	17.66	47.96	19.32	18.74	18.66	18.66	49.24
Aligned Rank Test																				
2	6.62	6.28	6.42	6.42	32.40	6.18	5.84	5.76	5.76	30.74	6.36	6.14	6.00	6.00	29.82	7.66	7.36	7.36	7.36	33.26
4	9.32	9.08	9.00	9.00	36.90	8.34	8.06	8.00	8.00	34.24	10.64	10.14	10.28	10.28	37.60	11.70	11.28	11.24	11.24	39.40
6	9.46	9.08	9.06	9.06	36.40	9.10	8.80	8.60	8.60	35.76	13.46	13.12	13.14	13.14	42.12	15.40	14.96	14.98	14.98	44.64
8	10.02	9.80	9.86	9.86	37.70	9.88	9.46	9.52	9.52	37.56	16.04	15.44	15.36	15.36	45.56	16.76	16.34	16.34	16.34	46.10
10	11.72	11.24	11.26	11.26	41.02	10.34	10.04	10.02	10.02	38.52	17.74	17.48	17.52	17.52	48.10	19.10	18.64	18.58	18.58	49.30
Sub-Sampling Method																				
2	5.64	5.20	5.12	5.12	30.94	5.20	4.82	4.86	4.86	29.10	4.90	4.72	4.62	4.62	29.62	5.36	5.02	5.04	5.04	30.74
4	6.36	5.90	5.90	5.90	33.46	5.80	5.74	5.66	5.66	31.20	6.54	6.32	6.24	6.24	32.54	6.18	5.82	5.82	5.82	33.16
6	5.28	5.10	5.18	5.18	31.28	5.56	5.32	5.24	5.24	31.34	7.06	6.82	6.64	6.64	32.70	6.92	6.62	6.72	6.72	33.14
8	5.60	5.32	5.26	5.26	32.64	6.02	5.90	5.82	5.82	31.40	7.12	6.92	6.82	6.82	34.24	7.42	6.96	7.12	7.12	32.94
10	5.68	5.56	5.62	5.62	34.48	6.10	5.92	5.86	5.86	32.54	7.96	7.78	7.76	7.76	33.46	7.94	7.56	7.54	7.54	33.40
MIXED model using the GROUP=option																				
2	5.62	5.24	5.40	5.40	30.36	5.10	4.92	4.76	4.76	29.50	5.52	5.52	5.34	5.34	30.40	6.44	5.97	6.32	6.32	31.68
4	5.52	5.29	5.38	5.38	32.88	5.68	5.40	5.44	5.44	30.92	7.32	7.24	7.18	7.18	33.60	7.36	6.84	7.00	7.00	32.86
6	5.22	4.92	5.92	4.92	30.88	5.19	5.02	4.92	4.92	30.52	7.95	7.84	7.56	7.56	33.76	7.98	7.53	7.58	7.58	33.26
8	5.14	4.96	5.08	5.08	32.74	5.28	4.98	4.86	4.86	32.02	8.09	7.98	7.86	7.86	34.26	8.12	7.83	7.82	7.82	33.30
10	5.23	4.97	5.02	5.02	34.44	5.36	5.24	5.12	5.12	32.08	8.59	8.43	8.40	8.40	33.72	8.53	8.21	8.28	8.28	33.86

Table 10.Methods for Type I error rates in Three Treatment Test under 5 Covariance Structures with $\sigma^2_{\gamma A}$ = 2; $\sigma^2{}_B$= (2, 4, 6, 8, 10);and σ^2_C=1.

$\sigma^2{}_B$/	Distribution Scenarios																			
	Z*w*~ Normal ***e*~** Skew					**Z*w*~** Normal ***e*~** Normal					**Z*w*~**Skew ***e*~**Normal					**Z*w*~** Skew ***e*~** Skew				
	AR(1)	TOEP	CS	UN	VS	AR(1)	TOEP	CS	UN	VS	AR(1)	TOEP	CS	UN	VS	AR(1)	TOEP	CS	UN	VS
Original Test																				
2	6.68	6.30	6.26	6.26	46.48	7.18	6.82	6.90	6.90	46.14	6.80	6.58	6.52	6.52	47.08	6.68	6.42	6.62	6.62	47.82
4	8.62	8.20	8.28	8.28	48.70	8.64	8.22	8.20	8.20	49.32	8.64	8.28	8.30	8.30	50.64	8.22	7.90	7.88	7.88	48.86
6	9.68	9.24	9.38	9.38	50.00	10.08	9.78	9.58	9.58	51.08	9.60	9.32	9.48	9.48	51.38	10.24	9.94	9.82	9.82	52.22
8	10.08	9.88	10.00	10.00	51.70	10.68	10.26	10.30	10.30	50.28	11.30	10.86	10.94	10.94	50.46	11.32	10.94	11.02	11.02	53.24
10	11.64	11.38	11.40	11.40	50.96	10.94	10.76	10.74	10.74	51.62	12.36	11.98	11.94	11.94	51.74	12.84	12.42	12.40	12.40	53.08
Sub-Sampling Method																				
2	4.96	4.52	4.48	4.48	43.12	5.50	5.24	5.36	5.36	42.28	5.18	4.86	4.84	4.84	43.06	5.04	4.66	4.68	4.68	44.40
4	5.20	4.88	4.98	4.98	42.58	5.84	5.58	5.48	5.48	43.58	5.88	5.62	5.60	5.60	45.78	5.50	5.26	5.38	5.38	43.48
6	5.40	5.06	5.24	5.24	42.66	6.50	6.14	6.16	6.16	44.30	6.26	6.14	6.02	6.02	44.32	7.04	6.68	6.72	6.72	44.16
8	6.16	5.98	6.04	6.04	43.42	6.48	6.28	6.24	6.24	43.04	7.06	6.80	6.68	6.68	42.76	7.24	7.06	7.14	7.14	44.84
10	6.38	6.28	6.40	6.40	42.44	6.20	6.04	6.04	6.04	41.92	7.76	7.48	7.50	7.50	43.60	8.42	8.12	8.12	8.12	44.78
MIXED model using the GROUP=option																				
2	4.90	4.53	4.76	4.76	43.92	5.76	5.06	5.18	5.18	43.66	6.47	6.02	5.96	5.96	44.76	6.72	6.09	6.28	6.28	45.56
4	5.50	5.15	5.12	5.12	44.44	5.80	5.49	5.22	5.22	44.40	7.18	6.77	6.74	6.74	47.44	7.46	6.98	7.14	7.14	45.68
6	4.82	4.59	4.64	4.64	45.38	5.86	5.49	5.38	5.38	45.92	7.17	6.87	6.54	6.54	47.00	7.88	7.38	7.52	7.52	47.28
8	5.20	4.66	4.76	4.76	45.88	5.69	5.67	5.52	5.52	45.36	8.01	7.70	7.54	7.54	46.74	8.50	7.82	8.00	8.00	48.24
10	4.69	4.23	4.40	4.40	45.94	5.19	4.81	5.00	4.98	45.48	8.33	7.97	7.74	7.74	46.42	8.63	8.35	8.46	8.46	48.30

5. CONCLUSIONand DISCUSSION

To sum up, this study simulated longitudinal data under four different conditions, then used the MIXED model to do analysis. The four conditions that were simulated include the following: 1. Unbalanced sample size; 2. Unequal group variance; 3. Violating the normality assumption of the MIXED model; and 4. A MIXED model with incorrect covariance structures. This research aims to check how Type I error rates would be affected within different conditions.

The first and main problem in this study is the analogue to the Behrens-Fisher problem under the MIXED model structure. There are two components to this problem: unbalanced group size (sample size ratio different than 1) and unequal group variance (variance ratio not equal to 1). When only one component assumption is violated, this is generally not a concern when using a MIXED model, but the Type I error rate will likely be inflated if the two component violations occur simultaneously. When the size and variance ratios are fixed, the inflated Type I error rate is consistent no matter what the actual sample sizes are or what the actual variances are. The number of repeated measures does not appear to affect the Type I error either. When a group has a small sample size in comparison to other groups, but a relatively large variance in comparison to other groups, we should be cautious of the Type I error inflation problem. The MIXED model method using the GROUP= option Method and Sub-Sampling Method can be reasonable solutions when having this problem. From a practical point, the method of the MIXED model using the GROUP= option is recommended.

The second problem in this study was in regards to violating the normality assumption of the MIXED model. The MIXED model is reasonably robust to modest violations of the normal distribution. However, when data is heavily skewed with a big difference between group variances, the MIXED model's performance will be not be as robust.

The third problem is how does the incorrect covariance structures affect Type I error rates. The Type I error rates were all compared when the correct covariance structure was First-Order Autoregressive (AR(1)) [3,6]. Choosing the incorrect covariance structure among Toeplitz (TOEP), Compound Symmetry (CS) and Unstructured (UN) does not affect the results of Type I error rates with the difference among the four being around one percent. Nevertheless, Variance Components (VC) structure [3,6] appears to increase the Type I error rate 3 to even 17 times compared to the results of First-Order Autoregressive (AR(1)). As we know, Variance Components (VC) is the simplest covariance structure. It specifies that observations are independent even on the same subjects, which is not realistic for most longitudinal data. So neglecting the correlated measurements in a longitudinal study might be why Variance Components (VC) structure causes the excessive Type I error rate inflation. When simulating AR(1) data in this study with the sample sizes used, we could not really distinguish a difference across the other methods. Future research would include simulating other types of data besides AR(1) and seeing if the results from the difference covariance structures (besides VC) are similar. We would also need to investigate different sample sizes.

Future research could also look into the impact of issues of unbalanced samples and heterogeneous variances in more complex designs. All of these simulations assume data from continuous distributions. The mixed model framework generalizes to discrete distributions as well. Perhaps some of these issues such as which variance-covariance structure to use or the impact of unequal variances and unequal sample sizes could be investigated using simulated data from discrete distributions. Thought would need to be given to the assumptions of these generalized mixed models and whether or not the sample size and variance issues that can plague continuous distributions in a mixed model ANOVA would impact the generalized models as well.

In addition to how the Type 1 error is affected, it is also important to consider the power of a test and how that is changed with various assumption violations. Researchers should have some idea as to the effect size that they would like to detect. Future research could include how the power of detecting a certain effect size is changed with violation of assumptions. This research would also include changes in sample sizes and sample size justification as in [31].

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

6. REFERENCES

[1]Guerin, L. and Stroup, W. W. (2000). A simulation study to evaluate procmixedanalysisofrepeatedmeasuresdata. Annual Conference on Applied Statistics in Agriculture – 12th Annual Conference Proceedings. New Prairie Press.

[2] Scheffe, H. (1959). *The analysis of variance*, volume 72. John Wiley & Sons.

[3] Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for mixed models, 2nd ed*. SAS Institute, Inc. Cary, NC.

[4] Kwok, O.-m., West, S. G., and Green, S. B. (2007). The impact of mis-specifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavioral Research*,42(3):557–592.

[5]SAS Institute, Inc. (2015). SAS/STAT. 14.2 User's Guide. *Cary, NC: SAS Institute Inc*.

[6] Kincaid, C. (2005). Guidelines for selecting the covariance structure in mixed model analysis. In *Proceedings of the thirtieth annual SAS users group international conference*, 30:198–130.SASInstituteIncCaryNC.

[7] Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated- measures models with structured covariance matrices. *Biometrics*, pages805–820.

[8] Keselman, H., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1999b). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*,52(1):63–78.

[9] Keselman, H., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1999a). The analysis of repeated measurements: A comparison of mixed-model Satterthwaite F tests and a nonpooled adjusted degrees of freedom multivariate test. *Communications in Statistics-Theory and Methods*,28(12):2967–2999.

[10] Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, pages 983–997.

[11] Prasad, N. N. and Rao, J. N. (1990). The estimation of the mean squared errorofsmall-areaestimators.*JournaloftheAmericanStatisticalAssociation*,85(409):163–171.

[12] Harville, D. A. and Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, 87(419):724–731.

[13] Fisher, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*,8(4):376–386.

[14] Kim, S.-H. and Cohen, A. S. (1998). On the Behrens-Fisher problem: A review. *Journal of Educational and Behavioral Statistics*,23(4):356–377.

[15] Paul, S., Wang, Y.-G., and Ullah, I. (2019). A review of the Behrens-Fisher problemandsomeofitsanalogs:Doesthesamesizefitall?*RevstatStatisticalJournal*,17(4):563–597.

[16] Hill, M. and Dixon, W. (1982). Robustness in real life: A study of clinical laboratory data. *Biometrics*, 38(2): 377–396.

[17] Reed III, J. F. (2003). Solutions to the Behrens–Fisher problem. *Computer methods and programs in biomedicine*, 70(3):259–263.

[18] Schielzeth H, Dingemanse NJ, Nakagawa S, Westneat DF, Allegue H, Teplitsky C, Réale D, Dochtermann NA, Garamszegi LZ, Araya-Ajoy YG. Robustness of linear mixed-effects models to violations of distributional assumptions. Methods in ecology and evolution. 2020 Sep;11(9):1141-52.

[19] Lehmann, E. L. and D'Abrera, H. J. (1975). *Nonparametrics: statistical methods based on ranks.*Holden-Day.

[20] Higgins, J. J., Blair, R. C., and Tashtoush, S. (1990). The aligned rank transform procedure. Conference on Applied Statistics in Agriculture – 2nd Annual Conference Proceedings. New Prairie Press.

[21]Hesterberg, T., Moore, D., Monaghan, S., Clipson, A., Epstein, R., Moore, D., and McCabe, G. (2005). Bootstrap methods and permutation tests. Introduction to the Practice of Statistics – Chapter 16. New York: W. H. Freeman.

[22] Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, 68(3):589–599.

[23] Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer Science & Business Media.

[24] Schroeder, W. J. and Martin, K. M. (2005). Overview of visualization. The Visualization Handbook, pp 3-35.

[25] Chernick, M. R. (2011). *Bootstrap methods: A guide for practitioners and researchers*. John Wiley & Sons.

[26] Wicklin, R. (2013). *Simulating data with SAS*. SASInstitute.

[27] Bradley,J.V.(1978).Robustness?*BritishJournalofMathematicalandStatistical Psychology*,31(2):144–152.

[28] Kowalchuk, R. K., Keselman, H., Algina, J., and Wolfinger, R. D. (2004). The analysis of repeated measurements with mixed-model adjusted f tests. *Educational and Psychological Measurement*,64(2):224–242.

[29] Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4):521–532.

[30] Vale, C. D. and Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometika,* 48(3): 465-471.

[31] Lakens D. Sample size justification. Collabra: Psychology. 2022 Mar 22;8(1):33267.

.