Research Article
AMERICAN JOURNAL OF PHARMACY AND HEALTH RESEARCH
www.ajphr.com
2014, Volume 2, Issue 8
ISSN: 2321–3647(online)

Statistic in Mixed procedure for Clinical Trials QTRS Studies

Vasanth Kumar Kunitala[1*], Abhilash Reddy Y[2], Purnachander K[3], ChandraShaker T[4], Nujamuddin Junaid K[5] , Jagruti D[6].
1.PAREXEL International, India, building no:-20, mid space, raheja IT park, Hyderabad.

ABSTRACT

As an very important aspect of the clinical evaluation of a phase-I trails is to administration of new drug then check the QT/QTc interval to determine the risk of cardiac repolarization Prolongation based on International Conference on Harmonization (ICH) E14 guidelines. To justify the safety of the patient is checking by the Electrocardiogram (ECG) parameters including PR and RR intervals, QTc Duration, QRS duration, T wave morphology, presence of U waves, and Outlier assessment. Hypothesis is Administration of a single dose of drug does not prolong the QTc interval to a clinically significant degree. Primary assessment of the true mean difference (drug-placebo) of QTc change from baseline is less than 10msec. We can estimate this Primary assessment by using more flexible and advance method proc mixed procedure. To identify a dose of the preliminary market formulation (PMF) of new drug that can achieve a safe and well tolerated maximum plasma concentration (Cmax) at least 5X higher than the Cmax associated with the clinical dose of new drug that can be known and justified by statistical assessment using Proc Mixed Procedure. Based upon this procedure we can estimate very import statistics like sample calculation, LS-means, CI, F-stats, Probability, chi-square etc. which can be justify the safe and tolerated of new drug.

Keywords: - Proc mixed, QT/QTc interval, LS-means, Probability, Benefit–risk assessment

*Corresponding Author Email vasanthmph@gmail.com
Received 10 July 2014, Accepted 23 July 2014

Please cite this article in press as: Kunitala V *et al*., Statistic in Mixed procedure for Clinical Trials QTRS Studies. American Journal of Pharmacy & Health Research 2014.

INTRODUCTION

In phase-I trails safety is very important this can also assess by QTRS studies. In QTRS study main importing thing is to observe changes in ECG parameters including PR and RR intervals, QTc Duration, QRS duration and T wave morphology, presence of U waves. Assess the true mean difference drug and placebo of QTc change from baseline is less than 10msec. An undesirable property of some non-antiarrhythmic drugs is their ability to delay cardiac repolarization leading to increased risk of development of life-threatening cardiac arrhythmias such as Torsade de Pointes (TdP), and possibly other ventricular tachyarrhythmia. A comprehensive evaluation of cardiac safety is required for the safe conduct of drug development programs and drug registrations. The potential effect of a drug on cardiac repolarization can be measured as prolongation of the QT interval on ECG recordings. There is in general a qualitative relationship between substantial QT prolongation and the risk of TdP. This proposed trial will evaluate in a rigorous manner the impact of administration of new drug on QT/QTc interval to determine the risk of cardiac repolarization Prolongation based on International Conference on Harmonization (ICH) E14 guidelines. We can estimate this Primary assessment by using proc mixed procedure. The mixed procedure fits a variety of mixed linear models to data and enables you to use these fitted models to make statistical inferences about the data. A mixed linear model is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and non-constant variability. The mixed linear model, therefore, provides you with the flexibility of modeling not only the means of your data but their variances and co-variances as well. The fixed-effects parameters are associated with known explanatory variables, as in the standard linear model. These variables can be either qualitative as in the traditional analysis of variance or quantitative as in standard linear regression. However, the covariance parameters are what distinguish the mixed linear model from the standard linear model. The need for covariance parameters arises quite frequently in applications, the following being the two most typical scenarios: 1) the experimental units on which the data are measured can be grouped into clusters, and the data from a common cluster are correlated. 2) Repeated measurements are taken on the same experimental unit, and these repeated measurements are correlated or exhibit variability that changes. Proc mixed provides a variety of covariance structures to handle the previous two scenarios. The most common of these structures arises from the use of random-effects parameters, which are additional unknown random variables assumed to impact the variability of the data. The variances of the random-

effects parameters, commonly known as variance components, become the covariance parameters for this particular structure. Traditional mixed linear models contain both fixed- and random-effects parameters, and, in fact, it is the combination of these two types of effects that led to the name mixed model. Proc mixed fits not only these traditional variance component models but numerous other covariance structures as well. Once a model has been fit to your data, you can use it to draw statistical inferences via both the fixed-effects and covariance parameters. Proc mixed computes several different statistics suitable for generating hypothesis tests and confidence intervals. The validity of these statistics depends upon the mean and variance-covariance model you select so it is important to choose the model carefully. Some of the output from proc mixed helps you assess your model and compare it with others. Proc mixed provides easy accessibility to numerous mixed linear models that are useful in many common statistical analyses. In the style of the GLM procedure, proc mixed fits the specified mixed linear model and produces appropriate statistics. Differences between the proc GLM and mixed was shown in table1.

Table 1 was shown the Differences between the proc GLM and Mixed.

Criteria	PROC GLM	PROC MIXED
Nutshell	Designed for **fixed effects** model with allowance for certain adjustments in the presence of random terms. For many designs special attention needs to be given to least square means and contrasts since their standard errors are not necessarily correct. This is true, for example, for split-plot designs.	Designed for **mixed effects** models. Random terms are incorporated into inference from the outset. Contrasts, least square means and estimates of linear combinations are reported with correct standard errors. Attention needs to be paid to the selection of an appropriate inference space.
Danger	If analysis is driven by accounting for degrees of freedoms and tests, p-values, contrasts, least square means etc. are taken for granted. Depending on the design, some results reported by PROC GLM are actually incorrect.	In order to access the powerful inferential features of PROC MIXED, effects are treated as random when they are actually fixed. Prime example: block effects.
Estimation of fixed effects	Based on Ordinary Least Squares	Based on Generalized Least Squares in a Gaussian error model (estimates are Maximum Likelihood Estimates under normality).
Estimation of variance	Method-of-Moments estimation	Maximum Likelihood or

components	(ANOVA method) of solving expected mean squares for the variance components	Restricted Maximum Likelihood estimation.
MODEL statement	All effects are listed, whether fixed or random	Only the fixed effects are listed. Random effects are listed in the RANDOM statement.
LSMEANS statement	Interprets all effects as **fixed**, even the random effects. While least square means are correct, their standard errors are not necessarily correct. This is true, even if a RANDOM statement is used.	Only effects in the MODEL statement are assumed fixed. The standard error estimates for least square means account for the random effects.
ESTIMATE statement	Same as LSMEANS statements	Same as LSMEANS statements
Sums of Squares	They are everywhere	Since it is not based on expected mean squares, there are no sums of squares in the standard output.
RANDOM statement	Invokes the calculation of expected mean squares for the listed effects and appropriate test using the /TEST option. The randomness of the effect is not incorporated into the tests of main effects, lsmeans, contrasts, etc.	Signals incorporation of the listed effects in all aspects of inference. Options allow selection of various correlation models describing the dependencies of multiple random effects.
TEST statement	Important statement to utilize correct error terms in testing model effects. Prime examples: Subsampling designs, split-plot designs	Gone.

Formulation of the Mixed Model

The previous general linear model is certainly a useful one and it is the one fitted by the GLM procedure. However, many times the distributional assumption about € is too restrictive. The mixed model extends the general linear model by allowing a more flexible specification of the covariance matrix of €. In other words, it allows for both correlation and heterogeneous variances, although you still assume normality.

The mixed model is written as: - $y = X\beta + Z\gamma + €$

Where everything is this equation same as in the general linear model except for the addition of the known design matrix Z and the vector of unknown random-effects parameters γ. The matrix Z can contain either continuous or dummy variables, just like X. The name mixed model comes from the fact that the model contains both of the fixed effects parameters, β, and random-effects parameters, γ. A key assumption in the foregoing analysis is that γ and € are normally distributed with

$$\mathrm{E}\begin{bmatrix}\gamma\\ \epsilon\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$$
$$\mathrm{Var}\begin{bmatrix}\gamma\\ \epsilon\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

The variance of y is, therefore, V = Z G Z' + R. You can model V by setting up the random-effects design matrix Z and by specifying covariance structures for G and R. Note that this is a general specification of the mixed model, in contrast to many texts and articles that discuss only simple random effects. Simple random effects are a special case of the general specification with Z containing dummy variables, G containing variance components in a diagonal structure and $\mathbf{R}= \sigma^2\mathbf{I_n}$, where In denotes the n ×n identity matrix. The general linear model is a further special case with Z = 0 and $\mathbf{R}= \sigma^2\mathbf{I_n}$. The following two examples illustrate the most common formulations of the general linear mixed model.

Example: - Growth Curve with Compound Symmetry

Suppose that you have three growth curve measurements for s individuals and that you want to fit an overall linear trend in time. Your X matrix is as follows:

$$\mathbf{X} = \begin{bmatrix}1 & 1\\ 1 & 2\\ 1 & 3\\ \vdots & \vdots\\ 1 & 1\\ 1 & 2\\ 1 & 3\end{bmatrix}$$

The first column coded entirely with 1s fits an intercept, and the second column coded with times of 1, 2, 3, fits a slope. Here n = 3s and p = 2. Suppose further that you want to introduce a common correlation among the observations from a single individual, with correlation being the same for all individuals. One way of setting this up in the general mixed model is to eliminate the Z and G matrices and let the R matrix be blocking diagonal with blocks corresponding to the individuals and with each block having the compound-symmetry structure. This structure has two unknown parameters, one modeling a common covariance and the other a residual variance. The form for R would then be as follows:

$$\mathbf{R} = \begin{bmatrix}\sigma_1^2+\sigma^2 & \sigma_1^2 & \sigma_1^2 & & & \\ \sigma_1^2 & \sigma_1^2+\sigma^2 & \sigma_1^2 & & & \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & & & \\ & & & \ddots & & \\ & & & & \sigma_1^2+\sigma^2 & \sigma_1^2 & \sigma_1^2\\ & & & & \sigma_1^2 & \sigma_1^2+\sigma^2 & \sigma_1^2\\ & & & & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2\end{bmatrix}$$

Where in blanks denote zeroes. There are 3s rows and columns altogether, and the common correlation is $\sigma_1^2/(\sigma_1^2+\sigma^2)$. The PROC MIXED code to fit this model is as follows:

```
proc mixed;
class indiv;
model y = time;
```

repeated / type=cs subject=indiv;

run;

Here, indiv is a classification variable indexing individuals. The MODEL statement fits a straight line for time; the intercept is fit by default just as in PROC GLM. The REPEATED statement models the R matrix: TYPE=CS specifies the compound symmetry structure, and SUBJECT=INDIV specifies the blocks of R. An alternative way of specifying the common intra-individual correlation is to let

$$\mathbf{Z} = \begin{bmatrix} 1 & & & \\ 1 & & & \\ 1 & & & \\ & 1 & & \\ & 1 & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & 1 \\ & & & 1 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_1^2 & & \\ & & \ddots & \\ & & & \sigma_1^2 \end{bmatrix}$$

And $\mathbf{R} = \sigma^2 \mathbf{I}_n$. The Z matrix has 3s rows and s columns, and G is s ×s. You can set up this model in PROC MIXED in two different but equivalent ways:

proc mixed;

class indiv;

model y = time;

random indiv;

run;

proc mixed;

class indiv;

model y = time;

random intercept / subject=indiv;

run;

Both of these specifications fit the same model as the previous one that used the REPEATED statement; however, the RANDOM specifications constrain the correlation to be positive whereas the REPEATED specification leaves the correlation unconstrained. Complete proc mixed syntax description was show in figure1.

```
PROC MIXED <options> method=type3/REML/ML/MIVQUE0;
        BY variables ;
        CLASS variables ;
        ID variables ;
        MODEL dependent = <fixed-effects> </ options> ;
        RANDOM random-effects </ options> ;
        REPEATED <repeated-effect></ options> ;
        PARMS (value-list) ...</ options> ;
        PRIOR <distribution ></ options> ;
        CONTRAST 'label' <fixed-effect values ...>
        <| random-effect values ...>, ...</ options> ;
        ESTIMATE 'label' <fixed-effect values ...>
        <| random-effect values ...></ options> ;
        LSMEANS fixed-effects </ options> ;
        LSMESTIMATE model-effect lsmestimate-specification < / options> ;
        SLICE model-effect </ options> ;
        STORE <OUT=>item-store-name </ LABEL='label'> ;
        WEIGHT variable ; RUN;

REML = residual (restricted) maximum likelihood,   ML = maximum likelihood,
MIVQUE0 =minimum variance quadratic unbiased estimation of the covariance parameters.
```

Figure 1 was shown the general syntax of the proc mixed procedure.

I) PROC MIXED statements:

In proc mixed statement the Items within angle brackets (< >) are optional. The CONTRAST, ESTIMATE, LSMEANS, MAKE, and RANDOM statements can appear multiple times; all other statements can appear only once. The proc mixed and MODEL statements are required, and the MODEL statement must appear after the CLASS statement if a CLASS statement is included. The CONTRAST, ESTIMATE, LSMEANS, RANDOM, and REPEATED statements must follow the MODEL statement. The CONTRAST and ESTIMATE statements must also follow any RANDOM statements. **II) BY Statement: -** You can specify a BY statement with proc mixed to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. **III) CLASS Statement: -** The CLASS statement names the classification variables to be used in the analysis. If the CLASS statement is used it must appear before the MODEL statement. Classification variables can be either character or numeric. The procedure uses only the first 16 characters of a character variable. Class levels are determined from the formatted values of the CLASS variables. Thus, you can use formats to group values into levels. You can adjust the display order of CLASS variable levels with the ORDER= option in the proc mixed statement. **IV) ID Statement: -** The ID statement specifies which variables from the input data set are to be included in the OUTP= and OUTPM= data sets from the MODEL statement. If you do not specify an ID statement then all variables are included in these data sets. Otherwise,

only the variables you list in the ID statement are included. Specifying an ID statement with no variables prevents any variables from being included in these data sets.

V) MODEL Statement: - MODEL dependent = < fixed-effects >< / options >; The MODEL statement determine the X matrix of the mixed model with a single dependent variable and the fixed effects. An intercept is included in the fixed-effects model by default. If no fixed effects are specified only this intercept term is fit. The intercept can be removed by using the NOINT option. You can specify the different options in the MODEL statement after a slash (/). **VI) MODEL Statement - SOLUTION Option: -** The SOLUTION option requests that a solution for the fixed-effects parameters be produced. The fixed-effects parameter estimates are $\hat{\beta}$ and their approximate standard errors are the square roots of the diagonal elements of $(X'\hat{V}^{-1}X)^{-}$. You can output this approximate variance matrix with the COVB option or modify it with the EMPIRICAL option in the proc mixed statement. Along with the estimates and their approximate standard errors, a t-statistic is computed as the estimate divided by its standard error. The degrees of freedom for this t-statistic match the one appearing in the "Tests of Fixed Effects" table under the effect containing the parameter. The "Pr > |t|" column contains the two-tailed p-value corresponding to the t-statistic and associated degrees of freedom. You can use the CL option to request confidence intervals for all of the parameters they are constructed around the estimate by using a radius of the standard error times a percentage point from the t-distribution. **VII) RANDOM Statement: -** RANDOM random-effects < / options >; The RANDOM statement defines the random effects constituting the γ vector in the mixed model. It can be used to specify traditional variance component models as in the VARCOMP procedure and to specify random coefficients. The random effects can be classification or continuous, and multiple RANDOM statements are possible. The purpose of the RANDOM statement is to define the Z matrix of the mixed model the random effects in the γ vector and the structure of G. The Z matrix is constructed exactly as the X matrix for the fixed effects and the G matrix is constructed to correspond with the effects constituting Z. The structure of G is defined by using the TYPE= option. You can specify INTERCEPT (or INT) as a random effect to indicate the intercept. Proc mixed does not include the intercept in the RANDOM statement by default as it does in the MODEL statement. You can specify the different options in the RANDOM statement after a slash (/). **VIII) REPEATED Statement: -** REPEATED < repeated-effect > < / options >; The REPEATED statement is used to specify the R matrix in the mixed model. If no REPEATED statement is specified R is assumed to be equal to. σ^2 I. For many repeated measures models no repeated effect is required in the REPEATED statement. Simply use the

SUBJECT= option to define the blocks of R and the TYPE= option to define their covariance structure. In this case the repeated measures data must be similarly ordered for each subject and you must indicate all missing response variables with periods in the input data set unless they all fall at the end of a subject's repeated response profile. These requirements are necessary in order to inform proc mixed of the proper location of the observed repeated responses. Specifying a repeated effect is useful when you do not want to indicate missing values with periods in the input data set. The repeated effect must contain only classification variables. Make sure that the levels of the repeated effect are different for each observation within a subject; otherwise, PROC MIXED constructs identical rows in R corresponding to the observations with the same level. This results in a singular R and an infinite likelihood. Whether you specify a REPEATED effect or not, in these rows of R for each subject are constructed in the order that they appear in the input data set. You can specify the following TYPE= option, in the REPEATED statement after a slash (/), few examples was shown in table 2.

Table 2 was shown the few options uses with type option in repeated statement.

Description	Structure	Example
Variance Components	VC (default)	$\begin{bmatrix} \sigma_B^2 & 0 & 0 & 0 \\ 0 & \sigma_B^2 & 0 & 0 \\ 0 & 0 & \sigma_{AB}^2 & 0 \\ 0 & 0 & 0 & \sigma_{AB}^2 \end{bmatrix}$
Compound Symmetry	CS	$\begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2+\sigma_1 \end{bmatrix}$
Unstructured	UN	$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$
Banded Main Diagonal	UN(1)	$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$

ix) REPEATED Statement - R Option: - R<=value-list> The R option requests that blocks of the estimated R matrix be displayed. The first block determined by the SUBJECT= effect is the default displayed block. Proc mixed displays blanks for value-lists that are 0. The value-list indicates the subjects for which blocks of R, are to be displayed. For example, Repeated / type=cs subject=person r=1, 3, 5. Displays block matrices for the first, third, and fifth persons. Each table created by PROC MIXED has a name associated with it and you must use this name to reference the table when using ODS statements. These names are listed in table 3.

Table 3 was listed few ODS statements used in proc mixed procedure.

Table Name	Description	Required Statement/ Option
AsyCorr	asymptotic correlation matrix of covariance parameters	PROC MIXED ASYCORR
Base	base densities used for posterior sampling	PRIOR
CholR	Cholesky root of blocks of the estimated **R** matrix	REPEATED / RC
CholV	Cholesky root of blocks of the estimated **V** matrix	RANDOM / VC
Contrasts	results from the CONTRAST statements	CONTRAST
ConvergenceStatus	convergence status	default
CovB	approximate covariance matrix of fixed-effects parameter estimates	MODEL / COVB
CovParms	estimated covariance parameters	default output
Diffs	differences of LS-means	LSMEANS / DIFF (or PDIFF)
Dimensions	dimensions of the model	default output
Estimates	results from ESTIMATE statements	ESTIMATE
FitStatistics	fit statistics	default
G	estimated **G** matrix	RANDOM / G
V	blocks of the estimated **V** matrix	RANDOM / V
LRT	likelihood ratio test	default output
LSMeans	LS-means	LSMEANS
MMEqSol	mixed model equations solution	PROC MIXED MMEQSOL
NObs	number of observations read and used	default output
ParmSearch	parameter search values	PARMS
RCorr	correlation matrix from blocks of the estimated **R** matrix	REPEATED / RCORR
Search	posterior density search table	PRIOR / PSEARCH
Slices	tests of LS-means slices	LSMEANS / SLICE=
SolutionF	fixed-effects solution vector	MODEL / S
SolutionR	random-effects solution vector	RANDOM / S
Tests1	Type 1 tests of fixed effects	MODEL / HTYPE=1
Type1	Type 1 analysis of variance	PROC MIXED METHOD=TYPE1
Trans	transformation of covariance parameters	PRIOR / PTRANS

The SUBJECT= option identifies the subjects in your mixed model. Complete independence is assumed across subjects; therefore, the SUBJECT= option produces a block-diagonal structure in R with identical blocks. When the SUBJECT= effect consists entirely of classification variables, the blocks of R correspond to observations sharing the same level of that effect. These blocks are sorted according to this effect as well. Continuous variables are permitted as arguments to the SUBJECT= option. Proc mixed does not sort by the values of the continuous variable rather it considers the data to be from a new subject or group whenever the value of the continuous variable changes from the previous observation. Using a continuous variable decreases execution

time for models with a large number of subjects or groups and also prevents the production of a large "Class Levels Information" table. If you want to model nonzero covariance among all of the observations in your SAS data set, specifies SUBJECT=INTERCEPT to treat the data as if they are all from one subject. Be aware though that, in this case, proc mixed manipulates an R matrix with dimensions equal to the number of observations. If no SUBJECT= effect is specified, then every observation is assumed to be from a different subject and R is assumed to be diagonal. For this reason, you usually want to use the SUBJECT= option in the REPEATED statement.

Methodology

In this proc mixed procedure following basic methods of statics are involved for estimation of data. Primary Estimation is done by pharmacokinetic parameters AUC0-24, C24hr and Cmax will be log transformed and analyzed based on a linear mixed effects model containing a fixed effect for treatment and a random effect for subject. The least squares means and corresponding 90% CIs for the log-transformed AUC0-24, C24hr and Cmax will be calculated from the model using the mean square error and referencing a t-distribution. The arithmetic mean on the log-scale and confidence limits will be exponentiation to obtain the AUC0-24, C24hr and Cmax geometric mean and 90% CIs for each level of treatment. The geometric means and their 90% CI will be tabulated for each treatment. T test is used to compare two different set of values. It is generally performed on a small set of data. T test is generally applied to normal distribution which has a small set of values. This test compares the mean of two samples. T test uses means and standard deviations of two samples to make a comparison. The formula for T test is given below:

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad \text{——————} \quad (1)$$

Where, x_1^- = Mean of first set of values, x_2^- = Mean of second set of values, S_1 = Standard deviation of first set of values, S_2 = Standard deviation of second set of values, n_1 = Total number of values in first set, n_2 = Total number of values in second set. Geometric mean = 1/n*(∑ log x), n = total number of observations, ∑ log x = sum of total observation's log base. GCV = Geometric coefficient of variation is calculated in the natural log-scale with the equation:-

GCV = 100 * sqrt(exp(s2) - 1) ———————— (2)

The variance of a group of values measures the spread of the distribution. A large variance indicates a wide range of values while a small variance indicates that the values lie close to their mean. The variance s^2 is calculated by summing the squared distances from each value to the mean of the values then dividing by one fewer than the number of observations. The standard deviation s is the square root of the variance

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 \quad \text{——————} \quad (3)$$

The sample mean, or average, of a group of values is calculated by taking the sum of all of the values and dividing by the total number of values. In other words for n values x1, x2, x3, ... , xn, the mean $\bar{x}$ = (x1 + x2 + x3 + ... + xn)/n, or

$$\bar{x} = \frac{1}{n}\sum x_i \quad \text{——————} \quad (4)$$

Confidence interval is a range within which most plausible values would occur. To calculate confidence interval, one needs to set confidence level as 90%, 95%, or 99% etc. Most commonly used confidence level is 95%. Confidence interval represents a particular interval within which the data is 95% (or whatever the confidence level chosen) sure or certain for a particular outcome. The formula for confidence interval is given below:

$$\text{If } n \leq 30, \text{ then } CI = \bar{x} \pm t_{\alpha/2} \frac{S}{\sqrt{n}} \quad \text{——————} \quad (5)$$

$$\text{If } n \geq 30, \text{ then } CI = \bar{x} \pm z_{\alpha/2} \frac{S}{\sqrt{n}} \quad \text{——————} \quad (6)$$

n = Number of terms, x = Sample Mean, σ = Standard Deviation, z α2 = Value corresponding to α2 in z table, t α2 = Value corresponding to α2 in t table α =1- confidence level100.The number of observation in a given sample population is known as Sample size. The sample size plays an important part in any study which helps us to find the difference between the populations from the given sample. Sample size can be smaller and larger, but the larger sample size gives us the more accurate results and in the lower case it is denoted by 'n' and the sample size in upper case is denoted by 'N'. The sample size formula for the infinite population is given as:

$$SS = \frac{Z^2 p(1-p)}{C^2} \quad \text{——————} \quad (7)$$

The sample size formula for the finite population is given as:

$$New\ SS = \frac{SS}{1 + (\frac{SS-1}{Pop})} \quad \text{(8)}$$

Here, SS = Sample size, Z = given z value, p = Percentage of population, C = Confidence level, Pop = Population, Z score is the outcome obtained when Z test is performed. Z test follows normal distribution under null hypothesis. Z score is calculated for a large number of data. To estimate Z score, we require a variable which is known as standardized random variable. This variable is denoted by x. We can find Z score when mean and standard deviation are known. We subtract the value of mean from standardized random variable and then divide the result by the value of standard deviation. The formula for calculating Z score is given below:

$$Z\ score = \frac{X - \bar{X}}{\sigma} \quad \text{(9)}$$

Where x = Standardized random variable, $\bar{x}$ = Mean, σ = Population standard deviation. Following is the formula for population standard deviation:

$$Population\ Standard\ Deviation = \sqrt{\frac{\sum_{i=1}^{n}(X - X_i)^2}{n}} \quad \text{(10)}$$

Where σ = Population standard deviation, xi = Items given $\bar{x}$ = Mean, n = Total number of items. Degrees of freedom are the number of values which are involved in the final calculation of a statistic that are expected to vary. In other words these are the independent part of data used in calculation. It is used to know the accuracy of the sample population used in research. The larger degree of freedom, is to larger the possibility of the entire population to be sampled accurately. The correct formula for degrees of freedom depends on the different situation. Here are few of them: for One sample t-test: DF = n-1 or Two sample t-test: DF =n1 + n2 – 2. Least-squares are one of the most commonly used methods in numerical computing. Essentially it is a technique for solving a set of equations where there are more equations than unknowns, i.e. an over determined set of equations. This set of notes shows the origins of a particular form of the algorithm batch linear least-squares. LS Mean = least squares mean, N denotes the number of subjects for each level of treatment used to estimate the LS means. Chi-square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. Chi square is one of the most useful of the non-parametric statistics. We use it when our data consist of people distributed across categories, and we want to know whether that distribution is different from what we would expect by chance or another set of expectations.

We don't have scores we don't have means. We just have numbers or frequencies. In other words we have nominal data.

The formula for chi square is:

$$X^2 = \sum \frac{(O - E)^2}{E} \quad \text{(11)}$$

Where X^2 is the value for chi square $\sum$ is the sum O is the observed frequency E is the expected frequency Chi-Square Statistic Suppose we select a random sample from a normal population. The chi-square statistic can be computed using the following equation: X2 = [(n - 1) * s2] / σ2, where n is the sample size, σ is the population standard deviation, s is the sample standard deviation equals and X2 is the chi-square statistic. F Test is a method to compare variance of two different set of values. F test is applied on F distribution under null hypothesis. For calculating F test value, we first find the mean of two given observations and then calculate their variance. F test value is expressed as the ratio of variances of two observations. The comparison between the variances of two sets of data can lead to many predictions. The formula for F test is given below

$$F\ value = \frac{variance\ 1}{variance\ 2} = \frac{\sigma_1^2}{\sigma_2^2} \quad \text{(12)}$$

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad \text{(13)}$$

Where, σ^2 = Variance, x = Values given in a set of data, $\bar{x}$ = Mean of the data, n = Total number of values. An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled. Exact "F-tests" mainly arise when the models have been fitted to the data using least squares. You can use the following equation to compute an f statistic: f = [s12/σ12] / [s22/σ22]. Where σ1 is the standard deviation of population 1, s1 is the standard deviation of the sample drawn from population 1, σ2 is the standard deviation of population 2 and s1 is the standard deviation of the sample drawn from population 2. The degree of freedom (v1) refers to the degrees of freedom associated with the sample standard deviation s1 in the numerator; and the degree of freedom (v2) refers to the degrees of freedom associated with the sample standard deviation s2 in the denominator. A cumulative probability is a sum of probabilities. In connection with the F distribution calculator, cumulative probability refers to the probability that

an f statistic will be less than or equal to a specified value. An f value also known as an f statistic is a random variable that has an F distribution. Here are the steps required to compute an

F value: 1) Select a random sample of size n1 from a normal population, having a standard deviation equal to σ1. 2) Select an independent random sample of size n2 from a normal population, having a standard deviation equal to σ2. 3) The f value is the ratio of s12/σ12 and s22/σ22. Thus f = [s12/σ12] / [s22/σ22]. 4) f = [X21 / v1] / [X22 / v2] or f = [X21 * v2] / [X22 * v1]. Every f statistic can be associated with a unique cumulative probability. This cumulative probability represents the likelihood that the f statistic is less than or equal to a specified value. A statistical measure of the dispersion of data points in a data series around the mean. It is calculated as follows:

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Expected Return}}$$

The represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other. Above mention each static can be calculate by proc mixed procedure. To make an example for 3- period cross over QTRS study treatment comparison at each time point can be done by following Proc Mixed SAS code :-

```
PROC MIXED data = data;                                    CLASS subject
period treatment time;                                     MODEL QTcF change = period
treatment time treatment x time / DDFM=KR; REPEATED subject x period / SUB=subject
TYPE=CSH;                                                  RANDOM subject ;
LSMEAN treatment time treatment x time / diff CL ALPHA=0.1 ;   ESTIMATE
'New Drug vs. Placebo' treatment 1 0 -1 / ALPHA=0.1 CL;    RUN;
```

Analysis of structure and confounding:-

Although the structure of the crossover design is conceptually quite simple but it is still hard to read the entire article due to varied terminology and notation. The confusion is coming from for the treatment combination. For that treatment combination some people call it sequence and some people call it group. Recall, in the design, a group of people take sequence 1 and the other group of people takes sequence 2. The sequence means a combination of treatment in specified order. In this discussion, it was denoted as Group. All the design factors are denoted as the following: treatment (TREATMENT), with levels "A" and "B", time period (PERIOD), with levels "1" and "2", Sequence group (GROUP), with levels "A then B" and "B then A". If crossover studies were full-factorial designs (with factors GROUP, TREATMENT, and

PERIOD), it could be possible to evaluate not only the main effects, but also the GROUP*TREATMENT, PERIOD*TREATMENT, GROUP*PERIOD, and PERIOD* TREATMENT*GROUP interactions. However, crossover studies are not full factorial designs. Not all combinations of factors appear in the study (there is no GROUP='A then B', PERIOD='1', TREATMENT='B' combination, for example). Also each estimate of a main effect is confounded with a two-factor interaction. For example, when the TREATMENT and PERIOD are fixed, then GROUP is fixed also, which means we cannot separate the main effect of group and the two-factor interaction of TREATMENT*PERIOD. That is if one of the main effects is significant it is impossible to tell whether the effect, its alias, or both are generating the significant result. One could argue that there is no reason to expect a significant affect involving GROUP because subjects are assigned to GROUPS at random. Therefore a significant GROUP effect should be interpreted as resulting from a PERIOD*TREATMENT interaction and not from a difference between GROUPS. For similar reasons, a significant PERIOD effect is not considered to be the result of a GROUP*TREATMENT interaction, nor is a significant TREATMENT effect considered to be the result of a GROUP*PERIOD interaction. Based on this references code and all above mention equations, the program was shown in figure 2.

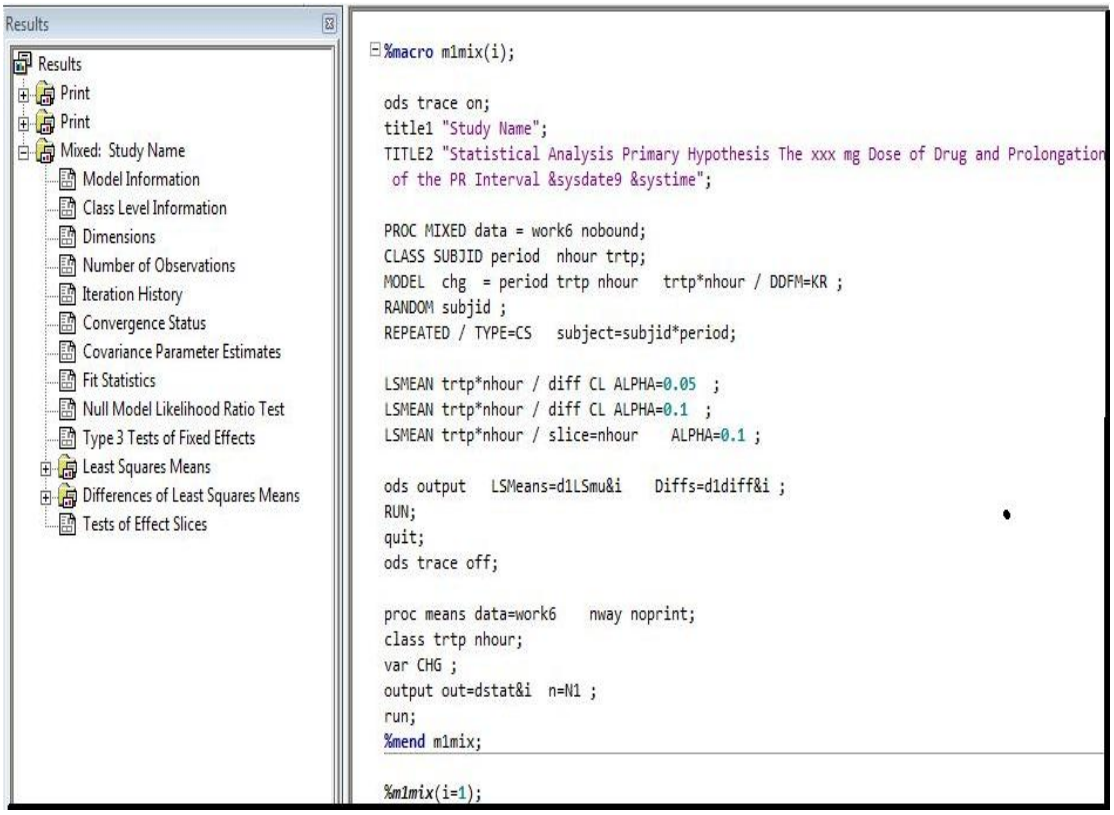

Figure 2 was shown the proc mixed procedure to analysis the primary end point for this study.

A few comments on the code: - CL (proc statement): This generates confidence limits for the variance-covariance estimates. DDFM=KR (model statement): The Kenward-Roger procedure is recommended for calculating the denominator degrees of freedom in repeated measures analyses. This procedure may result in denominator degrees of freedom that are not integers. R (repeated statement): This prints the covariance matrix. RCORR (repeated statement): This prints the correlation matrix. Proc mixed output data set was shown in table 4.

Table 4 was shown the output dataset of proc mixed procedure.

Effect	nhour	trtp	Estimate	StdErr	DF	tValue	Probt	Alpha	Lower	Upper
nhour*trtp	0.5	A	-0.486998	1.033726	397.3146	-0.47111	0.637821	0.05	-2.51925	1.545258
nhour*trtp	0.5	B	-1.378522	1.049161	395.4045	-1.31393	0.189632	0.05	-3.44115	0.684109
nhour*trtp	0.5	C	0.679942	1.047105	397.7553	0.649354	0.516484	0.05	-1.37861	2.738495
nhour*trtp	1	A	0.128386	1.033726	397.3146	0.124198	0.901222	0.05	-1.90387	2.160643
nhour*trtp	1	B	-2.378522	1.049161	395.4045	-2.26707	0.023925	0.05	-4.44115	-0.31589
nhour*trtp	1	C	-0.267426	1.047105	397.7553	-0.2554	0.79855	0.05	-2.32598	1.791127
nhour*trtp	1.5	A	-1.128024	1.033726	397.3146	-1.09122	0.275837	0.05	-3.16028	0.904233
nhour*trtp	1.5	B	-1.45747	1.049161	395.4045	-1.38918	0.165561	0.05	-3.5201	0.605161
nhour*trtp	1.5	C	1.574679	1.047105	397.7553	1.50384	0.133416	0.05	-0.48387	3.633232

Common proc mixed warning messages in log window:-

The warning **"**MIVQUE0 estimate of profiled variance is linearly related to other covariance parameters" also it was indicate that your model might be over specified. Try to re-specify your model so it is appropriate for your data. Select the *MIVQUE0* option button to estimate the variance components of the random effects in the model by Minimum Variance Quadratic Unbiased Estimators. METHOD= MIVQUE0 Procedure unbiased estimates that are invariant with respect to the fixed effects of model and are locally best quadratic unbiased estimates given to the true ratio of each component residual error component is zero. An alternative to ANOVA estimation is provided by maximum likelihood estimation. Maximum likelihood methods for estimating variance components are based on quadratic forms, and typically, but not always, require iteration to find a solution. Perhaps the simplest form of maximum likelihood estimation is MIVQUE(0) estimation. MIVQUE(0) produces Minimum Variance Quadratic Unbiased Estimators (i.e., MIVQUE). In MIVQUE(0) estimation, there is no weighting of the random effects (thus the 0 [zero] after MIVQUE), so an iterative solution for estimating variance components is not required. MIVQUE(0) estimation begins by constructing the Quadratic sums of squares (SSQ) matrix. The elements for the random effects in the SSQ matrix can most simply be described as the sums of squares of the sums of squares and cross products for each random effect in the model (after residualization on the fixed effects). The elements of this matrix provide coefficients, similar to the elements of the Expected Mean Squares matrix, which are

used to estimate the covariance's among the random factors and the dependent variable. The METHOD=TYPEn specifications apply only to variance component models with no SUBJECT= effects and no REPEATED statement. An analysis of variance table is included in the output, and the expected mean squares are used to estimate the variance components. The resulting method-of-moment variance component estimates are used in subsequent calculations, including standard errors computed from ESTIMATE and LSMEANS statements. For ODS purposes, the new table names are "Type1," "Type2," and "Type3," respectively. The METHOD=TYPE3 option in the PROC MIXED statement specifies to the procedure how to estimate variance component models (covariance parameters) using expected mean squares. The DDFM option in the MODEL statement specifies how the denominator degrees of freedom are calculated for estimating significance probabilities for the fixed effects.

Advantages of proc mixed procedure:-

1) Proc mixed syntax is different from that of the REPEATED statement in PROC GLM.
2) The specification of fixed effects is the same as in the GLM procedure; however, unlike PROC GLM you do not specify random effects in the MODEL statement. The MODEL statement is required.
3) GLM-type grammar, using MODEL, RANDOM, and REPEATED statements for model specification and CONTRAST, ESTIMATE, and LSMEANS statements for inferences
4) Appropriate standard errors for all specified estimable linear combinations of fixed and random effects, and corresponding t- and F-tests
5) Subject and group effects that enable blocking and heterogeneity, respectively
6) REML and ML estimation methods implemented with a Newton-Raphson algorithm
7) Capacity to handle unbalanced data
8) Ability to create a SAS data set corresponding to any table
9) Are used to predict the values of a numeric dependent variable
10) Assume the dependent variable is normally distributed

RESULT AND DISCUSSION:-

Proc mixed provides easy accessibility to numerous mixed linear models that are useful in many common statistical analyses. In the style of the GLM procedure, proc mixed fits the specified mixed linear model and produces appropriate statistics. proc mixed procedure we will get the complete statistical information regards the data like 1) model information 2) class information 3) dimensions 4) number of observations 5) iteration history 6) convergence status 7) covariance

parameter estimates 8) fit statistics 9) null model likelihood ratio test 10) type 3 tests of fixed effect. Proc mixed result output was show in figure 3.

The Mixed Procedure

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
SUBJID		3.8933
CS	SUBJID*PERIOD	5.1183
Residual		20.9220

Fit Statistics

-2 Res Log Likelihood	7530.9
AIC (smaller is better)	7536.9
AICC (smaller is better)	7536.9
BIC (smaller is better)	7542.1

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
2	252.04	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
PERIOD	2	70.5	1.08	0.3444
trtp	2	70.5	66.71	<.0001
nhour	10	1118	67.29	<.0001
nhour*trtp	20	1118	3.47	<.0001

Least Squares Means

Effect	trtp	nhour	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
nhour*trtp	A	0.5	-2.6954	0.8752	551	-3.08	0.0022	0.05	-4.4146	-0.9762

Figure 3 was shown the proc mixed procedure result output.

A repeated measures mixed model appropriate for a 3-period crossover design will be used to analyze Holter QTcF change from baseline. The model will include fixed factors period, treatment, time, treatment-by-time interaction, and a heterogeneous compound symmetry covariance structure will be assumed. The within-subject correlation across periods will be modeled by specifying subject as a random effect. The within-subject correlation across time points within a period will be modeled by specifying the subject by period interaction as the repeated measure with residual compound symmetry. Carryover will be investigated and tested at alpha=0.10, and will be included in the final model if found to be statistically significant. All other tests will be performed at alpha=0.05. To address the primary hypothesis that administration of a single xxx mg dose of drug does not prolong the QTcF interval to a clinically significant degree, means and two-sided 90% CIs, equivalent to one-sided upper 95% CI for the true mean difference drug placebo in QTcF change from baseline will be provided at each pre

specified Holter ECG time point. These CIs will be calculated using the appropriate error term from the model and referencing a t-distribution. If all of the upper limits of the 90% confidence intervals fall below 10 msec, the hypothesis that administration of a single xxx mg dose of drug does not prolong the QTc interval to a clinically significant degree will be supported. However, if any of the upper limits exceed 10 msec, then there will be insufficient evidence to accept the primary hypothesis. The comparison of drug to placebo will be conducted at each post-dose Holter ECG extraction time points: 0.5, 1, 1.5, 2, 3, 4, 6, 8, 12, 16, and 24 hours. According the time points ECG parameters changes in new drug vs references drug and placebo was shown in table 5.

Table 5 was shown the different time points ECG parameters changes in new drug vs references drug and placebo.

New Drug Prolongation of the QTcF Interval

			QTcF Value (msec)		Change from Baseline (msec)		Difference from Placebo (msec)	
Treatment	Time (h)	N	LSMean	95% CI	LSMean	95% CI	LSMean Difference	90% CI
Placebo	0.5	38	403	(398, 408)	-2.22	(-3.97, -0.489)		
	1	38	403	(397, 408)	-2.22	(-3.97, -0.489)		
	1.5	38	402	(397, 407)	-2.78	(-4.52, -1.04)		
	2	38	403	(398, 408)	-1.75	(-3.49, -0.0155)		
	3	38	404	(399, 409)	-0.992	(-2.73, 0.748)		
	4	38	405	(400, 410)	-0.150	(-1.89, 1.59)		
	6	37	394	(389, 399)	-10.6	(-12.4, -8.90)		
	8	38	397	(392, 402)	-7.38	(-9.12, -5.64)		
	12	38	398	(393, 403)	-6.51	(-8.25, -4.77)		
	16	38	404	(399, 409)	-0.387	(-2.12, 1.35)		
	24	38	401	(396, 406)	-3.38	(-5.12, -1.64)		
New Drug	0.5	39	402	(397, 407)	-2.69	(-4.41, -0.976)	-0.465	(-2.38, 1.45)
	1	39	403	(398, 408)	-1.43	(-3.15, 0.280)	0.791	(-1.12, 2.71)
	1.5	39	403	(398, 408)	-1.92	(-3.64, -0.207)	0.856	(-1.06, 2.78)
	2	39	403	(398, 408)	-1.74	(-3.46, -0.0275)	0.00930	(-1.91, 1.93)
	3	39	404	(399, 409)	-1.00	(-2.72, 0.716)	-0.0103	(-1.93, 1.91)
	4	39	404	(399, 409)	-0.336	(-2.05, 1.38)	-0.185	(-2.10, 1.73)
	6	39	394	(389, 399)	-10.8	(-12.5, -9.13)	-0.183	(-2.11, 1.75)
	8	39	397	(392, 402)	-7.90	(-9.61, -6.18)	-0.513	(-2.43, 1.41)
	12	39	398	(393, 403)	-6.92	(-8.64, -5.20)	-0.407	(-2.32, 1.51)
	16	39	404	(399, 409)	-0.874	(-2.59, 0.844)	-0.487	(-2.40, 1.43)
	24	39	402	(397, 407)	-2.97	(-4.69, -1.25)	0.410	(-1.50, 2.33)
References	0.5	38	406	(401, 411)	-0.0220	(-1.76, 1.72)	2.21	(0.273, 4.14)
	1	38	413	(407, 418)	6.35	(4.60, 8.09)	8.58	(6.64, 10.5)
	1.5	38	411	(406, 416)	4.74	(3.00, 6.48)	7.52	(5.59, 9.46)
	2	38	413	(408, 418)	6.82	(5.08, 8.56)	8.58	(6.64, 10.5)
	3	38	414	(409, 419)	7.61	(5.87, 9.35)	8.60	(6.67, 10.5)
	4	38	414	(409, 419)	7.85	(6.10, 9.59)	8.00	(6.06, 9.93)
	6	38	403	(398, 408)	-3.33	(-5.08, -1.59)	7.33	(5.38, 9.27)

This analysis will also be used to report the least squares means and 95% CIs for the change from baseline in QTcF by treatment and time point. Means and two-sided 90% CIs for the differences between each drug dose and placebo in QTcF change from baseline may be provided for other time points that are of clinical interest. A repeated measures mixed model with fixed factors period, treatment, time, and treatment-by-time interaction was used. The analysis assumed a compound symmetric covariance structure to model the within-subject correlation over time. Dependent variable in the model was the Holter QTcF change from baseline. If all of

the upper limits of the 90% confidence intervals for the mean difference fall below 10 msec, suggests the new dose does not prolong the QTcF interval. If at least one upper limit of the CI is above 10 msec, possible evidence of QTcF prolongation.

CONCLUSION

In QTRS studies to describe changes in electrocardiogram (ECG) parameters including PR and RR intervals, QRS duration, T wave morphology, presence of U waves, and outlier assessment is very important for new drugs in clinical studies. Proc mixed procedure is extraordinarily flexible and very use efficient procedure for estimating primary and secondary hypotheses. To identify a dose of the preliminary market formulation of new drug that can achieve a safe and well tolerated maximum plasma concentration (Cmax) at least 5X higher than the Cmax associated with the clinical dose of new drug that can be known and justified by statistical assessment using Proc Mixed Procedure. Based upon this procedure we can estimate very import statistics like sample calculation, LS-means, CI, F-stats, Probability, chi-square etc. which can justify the safe and tolerated of new drug.

REFERENCES

1) Galecki, A.T. Communications in Statistics Theory and Methods, 1994; 23(11), 3105–3119.
2) Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M., Journal of the American Statistical Association, 1990; 85, 972–985.
3) Ghosh, M., Schervish, M., “Bayesian Analysis of Linear Models,” 1992;432–433.
4) Giesbrecht, F.G.Applications of Mixed Models in Agriculture and Related Disciplines,1989;183–201
5) Giesbrecht, F.G. and Burns, J.C. (1985), “Two-Stage Analysis Based on a Mixed Model: Large-sample Asymptotic Theory and Small-Sample Simulation Results,”Biometrics, 1985;41, 477–486.
6) Pal SK. Complementary and alternative medicine: An overview. Current Science 2002.82, 518-24
7) Hanks, R.J., Sisson, D.V., Hurst, R.L, and Hubbard K.G., Soil Science Society American Journal, 1980; 44, 886–888.
8) Hannan, E.J. and Quinn, B.G. Journal of the Royal Statistical Society, Series B, 1979; 41, 190–195.
9) Dempster, A.P., Laird, N.M., and Rubin, D.B., Journal of the Royal Statistical Society, Ser. B., 1977; 39, 1–38.

zetapotenti
al

10) Goodnight, J.H., American Statistician,1979; 33, 149–158.
11) Fuller, W.A. and Battese, G.E., Journal of the American Statistical Association, 1973; 68, 626–632.
12) Fai, A.H.T. and Cornelius, P.L., Journal of Statistical Computation and Simulation, 1996; 54, 363–378.
13) Akaike, H. IEEE Transaction on Automatic Control 1974; AC–19, 716–723.