Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Power calculations for complex designed clinical trials using linear mixed models

Promotor :
dr. Francesca SOLMI

Promotor :
Dr. DAN LIN

Jedelyn Cabrieto

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

2013•2014

FACULTY OF SCIENCES

Master of Statistics

Master's thesis

Power calculations for complex designed clinical trials using linear mixed models

Promotor :
dr. Francesca SOLMI

Promotor :
Dr. DAN LIN

Jedelyn Cabrieto

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Acknowledgements

Though I know they could not really measure up to what I owe you, I wish to say my thanks to all who have helped me while I was doing the Biostatistics Master's program, especially in finishing this thesis.

To Dan Lin, Ph. D., for entrusting me this topic, and for providing me with insights and the needed encouragements, my sincerest gratitude to you. I have learned so much both in theory and in application through you and the discussions with your colleagues. I also want to thank them and Zoetis for choosing me to work on this project. To Francesca Solmi, Ph. D., my great appreciation for your theoretical suggestions and for all the "going out of your way" efforts to help me during code and report writing.

I would also like to thank all my CenStat professors who taught us well and challenged us to always think critically. To my classmates, from whom I learned so much, both from our classes and in our random lunch or coffee break conversations about our own little (some large) countries, thank you. I am also grateful to Mrs. Martine Machiels who has always been there to help us. To Adriana, Stellah, Thao and Olina, thank you for the friendship. And Marijke and Ewoud, I really appreciate all the help especially during first year when I was just learning Statistics again. To Kevin, Mohammed and Lazaro, second year masters was painfully hard especially at the end. But you were still as efficient, as dedicated and as cheerful as our fist meeting. I am really lucky to have worked with the best!

To Ate Chella, Cris, John, Ate Rochelle, Kuya Johan, Nay Gemma, Nong Guido, Cesar and Nolen, thank you for providing me with another home here in Belgium. I need not say more. To Ma'am Tina, I am greatly thankful you are always there when I need advice and encouragement. To Glenn and to the ladies from Kopierwiek, thank you for rescuing me during my stressed days.

I also would like to express my sincerest gratitude to VLIR-UOS, my scholarship sponsor, for giving the opportunity and providing financial support which enabled me to pursue a masters degree here in Belgium. To Prof. Geraldine Garcia and to Nolen, for encouraging me to apply for the scholarship, and to Prof. Formacion, Prof. Balinas and Prof. Faina for helping me with the application, thank you.

Finally, to Nanay and Tatay and to the fun people I grew up with at home - Raquel, Jerald, Jaide, Jeneil, Judy Ann, Jimmy and Julius, salamat! From you, I have learned to love asking questions and to dream of answers. I know I do not have to achieve anything for you to be proud of me, but this one I was able to finish because of thoughts of you.

And to God and all unnamed people who have helped me along my way, the biggest thanks are yours.

Jedelyn Cabrieto
Diepenbeek, 10 September, 2014

POWER CALCULATIONS FOR COMPLEX DESIGNED CLINICAL TRIALS USING LINEAR MIXED MODELS

by

Jedelyn Cabrieto

Hasselt University, 2014
Under the Supervision of
Dan Lin, Ph. D.
and
Francesca Solmi, Ph, D.

Abstract

Power calculation is a crucial part of planning a clinical trial to ensure that it is capable of detecting a clinically and statistically significant treatment difference. Complex designed veterinary clinical trials considered in this report have structures that could be naturally handled by linear mixed models by accounting for different sources of variation through the inclusion of random effects. However, definitive formulations for power calculations using linear mixed models do not exist for most cases. Thus, the primary aim of the investigator is to develop SAS macros that would generate data according to common experimental settings, and make power calculation possible for linear mixed models employed through extensive simulations. Superiority testing was done through approximate F-test for fixed effects in Proc Mixed and Proc Glimmix for continuous and binary data, respectively. For non-inferiority testing of continuous data, approximate t-test confidence interval was constructed around the treatment difference and was compared to the clinically acceptable margin. However, for non-inferiority testing of binary data, the clinically acceptable margin of difference is usually expressed in difference of proportions or odds ratio, while the confidence interval for treatment difference constructed by SAS is in the logit scale. Three methods then were proposed in order to conduct non-inferiority testing in this case, which were constructing the CI for difference of proportions (Independence), CI for difference of proportions (Delta Method) and CI for odds ratio. The SAS macros calculated power estimates coherent with specified parameters and experimental designs. In addition, they monitored convergence rate to provide a measure for the reliability of the power estimates generated.

Keywords: *Power; Clinical Trial; Linear Mixed Models; Superiority Testing; Non-Inferiority Testing; Approximate F-test, Approximate t-test, Confidence Interval, Difference of Proportions; Delta Method; Odds Ratio*

Dan Lin, Ph. D.

10 September, 2014

Francesca Solmi, Ph. D.

Contents

List of Tables

List of Figures

1 Introduction

New drugs and treatments are successfully introduced by pharmaceutical companies in the market when they are found to be more efficacious than existing standard treatments, or shown to be equally effective, but are easier to administer, less costly, have fewer side effects, or have more practical advantages contributing to better treatment results. Thus, clinical trials have been an indispensable tool for drug developers to exhibit superiority of a new treatment by showing a significant treatment difference over the standard treatment. They have also become the established way to show non-inferiority of an experimental drug when the treatment difference lies within a pre-specified clinically acceptable margin [17].

Exhibiting superiority and non-inferiority involve statistical tests which could only be reliable when their power, which is the probability of detecting a desired magnitude of existing treatment difference, is sufficient. Experimental designs then should be carefully drafted such that the number of subjects to be included in the trial would correspond to an acceptable power. Otherwise, the conduct of the experiment is futile simply because the trial itself is not powerful enough to detect the difference even if it exists [15]. In addition, there would be a serious loss of resources and grave ethical consequences of exposing subjects to non-standard treatments with no assurance that the study will gain useful medical knowledge. Thus, power calculation is a crucial part of a good clinical trial design, and it is included in the guidelines to be followed for approval of new drugs, treatments and diagnostic procedures [14].

This report tackled power calculation for complex designed clinical trials. The settings were specifically suited for veterinary clinical trials where experimental designs used blocking to control for known sources of variation among animals, and grouping them in pens were done when treatments or feeds could not be administered individually [5]. There were sixteen settings considered depending on the location of trial, whether it was done in a single center or in multiple centers, on the experimental unit, whether it was the animal or the pen, on the type of trial, whether it was for superiority or non-inferiority, and on the type of response, whether it was continuous or binary. For each setting, three experimental blocking designs were looked at, whether it was done through Complete Randomized Design (CRD), Randomized Complete Blocked Design (RCBD) or Generalized Randomized Blocked Design (GRBD).

For these settings, linear mixed models were employed as they naturally describe complex data structures, which could not be handled by fixed effects models [7]. Center, center by treatment interaction, block and pen effects were analyzed as random effects so that conclusions could be generalized to a broader inference space. However, existing power formulations only deal with fixed effects models where the distribution of the test statistic under the alternative hypothesis is known[18]. For Randomized Clinical Trials (RCTs), deterministic formulas in calculating sample sizes assuming classical models are well documented in the literature [12] . This is not the case for mixed models where the test statistic distribution is only known under the null hypothesis. Recent approaches to make this power calculation possible is by analytical approximation of the non-central F-distribution of the test statistic under the alternative hypothesis, and alternatively, by direct computation of power through extensive simulations [24]. While the first approach is relatively faster since it uses an ideal data set, it is not as comprehensive nor as accurate compared to the latter method[18].

In this project, SAS macros were developed to answer that need of having a tool to calculate power of experiments with settings described above. It employed the second approach, which was conducting extensive simulations to definitively compute power of linear mixed models with varying parameters and experimental designs. Additionally, the macros also monitored convergence rates of fitted linear mixed models with the objective of checking if power estimates generated were reliable such that they were based on a large number of converged models. It could also help planners determine which experimental designs and settings would pose future convergence issues, which could be a considerable difficulty especially in binary data analysis[19].

For superiority trials, both for continuous and binary responses, determining the significance of a treatment difference was done by looking at the approximate F-test for fixed effects in mixed models[7]. For non-inferiority trials where confidence intervals were the basis of testing [5], the procedure was straightforward, only for experiments with continuous responses. SAS provided outputs for confidence limits of the treatment difference using the approximate t-distribution under null, and the lower limit was compared with the pre-specified margin. For binary outcomes, however, estimation and test for treatment difference was conducted by SAS in the logit scale. Non-inferiority testing then for this case was not straightforward as most non-inferiority tests on binary outcomes were done on the difference of proportions or odds ratio[23]. Hence, several methods were proposed in this report to conduct non-inferiority tests for binary outcomes. The first method was by constructing the confidence interval for difference of proportions using its standard error computed assuming independence. The second approach was by constructing the same confidence interval using the standard error approximated through delta method. And finally, the third method was by constructing the confidence interval for odds ratio.

Results showed that for simple structured designs with minimal random effects included in the model, tests using the CI for difference of proportions (Independence) and the CI for difference of proportions (Delta Method) generated similar results. But for the most complex simulated trials with multiple center and pen as the experimental unit, CI for difference of proportions (Independence) proved to be extremely conservative than the other two methods, giving considerably low power estimates. CI for the odds ratio was consistently conservative, but it was able to take into account the features of the design compared to the previously mentioned method. On the other hand, the CI for difference of proportions (Delta Method) generated the highest power estimates, and proved to be flexible in accounting for the complexity of the experimental designs considered.

The SAS macros generated results which were expected for power calculations with respect to the specified model parameters and the experimental designs. From the simulations, blocking increased the power of an experiment when there was a considerable block variability. Generally, increasing the number of animals resulted to higher power. However, it was not always the case when pen was the experimental unit, wherein the number of animals per pen would reach a certain threshold for power maximally achievable for a certain experimental design. Finally, multi-center trials had more power when there were more centers included, and when center by treatment interaction variability was minimal.

2 Methodology

2.1 Background on Experimental Settings and Factors

Veterinary clinical trials are conducted in a variety of settings. Listed below were the experimental design factors which determined the types of settings considered in this report . Consequently, they were the basis of the data structure for the simulations and of the models fitted. Descriptions for items *a* and *b* were taken mainly from *Guideline on Statistical Principles for Clinical Trials for Veterinary Medicinal Products (Pharmaceuticals)* [5].

a. Single vs Multi-Center Locations
   In the current practice of veterinary clinical trials, the major aim of developing a new drug is to determine its dose or dose range which will be optimally effective and reliably safe for the target species. There are two major types of trials depending on their aims, the first of which is exploratory or pilot studies, and the next one, confirmatory, which usually concerns on dose determination, dose confirmation and field controlled studies, wherein the new drug is compared to a placebo or a standard treatment.

   Generally, exploratory trials are done in a single center. But there are also confirmatory trials conducted in a single location. Having trials with multiple centers, however, could be more preferred for two main reasons. First, it is an accepted way of evaluating a new medication more efficiently. Subjects are easily accrued when there are several sites included in the study, and that makes the conduct of the trial feasible for a given time-frame. Second, the generalization of conclusions from the study could be applied to different clinical settings, investigators, and geographical locations. Multicenter trials provide a setting closer to that of the actual scenario when the drug will be used in the future, wherein it will be administered by different medical personnel with varying expertise, or it will be given in different areas with varying environmental conditions, or for some other reasons which may influence its effect because of location.

b. Animal vs Pen as Experimental Units
   The experimental unit is the smallest unit in the experiment to which the treatment is independently applied[21]. Since investigational veterinary drugs are usually targeted to individual animals, it follows that animals are used as the experimental unit of trials. However, there are cases when they are housed together in group such as dogs in kennels, chickens in pens, or fish in tanks. When the animals cannot receive treatment or feeds individually, then the housing unit is used as the experimental unit. However, this is not only applicable for housing units but also in cases where the response is taken from a subunit of an animal like udder quarter of cows, or also when animals can be grouped according to biological factors like belonging in the same litter.

c. Continuous vs Binary Responses
   Clinical trials are conducted to answer a specific objective, and this is done through the analysis

of the primary endpoint, which is the response variable that is equally sensitive and clinically relevant. The statistical analysis would depend on the objective of the trial, whether it is conducted to exhibit efficacy, safety or both. But the type of primary endpoints also influences the analysis. Continuous responses are quantitative responses, examples of which are weight, bacterial count in milk or litter size. Binary responses are dichotomized responses such as cured or not cured, seropositive or not, or dead or alive. Dichotomization results to loss of efficiency[22] , and consequently loss of power. Thus, experiments with binary responses would require more sample sizes to achieve a certain power, compared to experiments with continuous responses.

d. Completely Randomized vs Blocked designs
   In experiments, there are certain factors which might influence the values of the response variables, but estimating them is not of interest to the investigator. These nuisance factors, when unknown, could be controlled through randomization [20]. When randomization is done in such a way that all experimental units have an equal chance of receiving the treatment, then the trial has a Completely Randomized Design (CRD). There are cases, however, when these nuisance factors are known and controlling for them is possible. Blocking then can be used as an important design technique [20], and randomization is done within the block, wherein experimental units are more homogeneous. Batches, position of pen in the lab or other baseline animal characteristics, are usual blocking factors in veterinary clinical trials. When there is only one animal (or pen when it is the experimental unit) per treatment within a block, it is referred to in this report as a Randomized Complete Blocked Design (RCBD). When there are two or more experimental units assigned to a treatment within a block, then it is referred to as Generalized Complete Block Designs (GRBD).

For the experimental settings considered, estimation of effects of center, center by treatment interaction, block and pen was not of interest. However, these effects should be accounted for to arrive at correct statistical inferences on the treatment effects. The two approaches possible for analyzing these effects were by treating them as either fixed or random[20]. Analyzing them as fixed effects would entail calculations of standard errors that would only account for one source of variation, which is the error term. Duchateau et. al[7] proposed that treating these effects as random would give the investigator the ability to apply conclusions to the desired inference space. Treating the block effect as a random effect, for instance, would allow one to calculate standard errors for treatment difference accounting for the variability between blocks. The conclusions generated could then be applied to the population of all blocks, which was not possible in the fixed effects model where conclusions were only valid to the specific blocks included in the study.

The flexibility inherent in mixed models to conduct the analysis in the most appropriate inference space and its capability of handling complex experimental designs[7] which was the case of veterinary experiments considered, made it an optimal model choice for this report. The following methodology would revolve on this modelling framework.

2.2 General Modelling Framework

2.2.1 General Linear Mixed Model

For continuous responses, the general form of the Linear Mixed Model employed was given by the following [7] ,

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \varepsilon \tag{1}$$

where,

$\mathbf{Y}$ =response vector

$\mathbf{X}$ =design matrix of fixed effects

β =vector of fixed effects

$\mathbf{Z}$ =design matrix of random effects

$\mathbf{b}$ =vector of random effects

ε =vector of residual terms

$\mathbf{b} \sim N(\mathbf{0}, \mathbf{D})$

$\varepsilon \sim N(\mathbf{0}, \Sigma)$

$\mathbf{b_1} \dots \mathbf{b_l}$, $\varepsilon_\mathbf{1} \dots \varepsilon_\mathbf{N}$ are independent.

Random effects included would vary for the different settings considered subsequently. However, the assumption that all random effects were independent and were drawn from a normal distribution with mean zero and a diagonal variance-covariance matrix would hold.

Estimation of fixed effects and their variances were discussed in detail in Duchateau et. al. (1997) where assuming the following distribution for the response vector $\mathbf{Y}$ for N subjects,

$$\mathbf{Y} \sim \mathbf{MVN}(\mathbf{X}\beta, \mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \sigma^\mathbf{2}\mathbf{I}_N)$$

the log-likelihood was given below,

$$l_\mathbf{Y}(\beta, \mathbf{V}) = -\frac{N}{2}log(2\pi) - \frac{1}{2}log|\mathbf{V}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$$

and when maximized with respect to β and set equal to $\mathbf{0}$ would give

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-\mathbf{1}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{V}^{-\mathbf{1}}\mathbf{Y}.$$

2.2.2 Generalized Linear Mixed Model

For binary reponses, the logit link is the canonical link function of the binomial distribution [1] and it naturally deals with the dichotomized nature of the data. Thus, the following Generalized Linear Mixed Model with logit link was employed.

$$logit(\pi) = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} \tag{2}$$

where,

$$\begin{aligned}
\mathbf{Y}|\mathbf{b} \sim Bernoulli(\pi) &= \text{response vector}\\
\mathbf{X} &= \text{design matrix of fixed effects}\\
\beta &= \text{vector of fixed effects}\\
\mathbf{Z} &= \text{design matrix of random effects}\\
\mathbf{b} &= \text{vector of random effects}\\
\mathbf{b} &\sim N(\mathbf{0},\mathbf{D})\\
\mathbf{b_1}\ldots\mathbf{b_l}, &\text{ are independent.}
\end{aligned}$$

The same assumptions on the random effects as with the continuous case were made here. However, the peculiarity in this model was that the random effects were plugged in the logit scale. Agresti (2006) noted that it is both convenient and natural in many applications when random effects enter the model on the same scale as the predictor scale. For instance, random effects may explain the variability caused by omitting certain explanatory variables or by other forms of missing data.

Molenberghs et. al. (2005) elaborated on how likelihoods of generalized linear mixed models could be approximated. For independent responses Y_{ij}, for instance, with vector of random effects $b_i \sim N(0,D)$, the density is given by

$$f(y_{ij}|\theta_{ij},\phi) = exp(\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij},\phi)),$$

with,

$$\begin{aligned}
\eta(\mu_{ij}) &= \eta[E(Y_{ij|b_i})] = x'_{ij}\beta + z'_{ij}b_i \ = \theta_{ij} \text{ for a known link function } \eta(.)\\
x_{ij} &= \text{p-dimensional vector of covariate values for fixed effects}\\
z_{ij} &= \text{q-dimensional vector of covariate values for random effects}\\
\beta &= \text{p-dimensional vector of fixed effects}\\
\phi &= \text{scale parameter}
\end{aligned}$$

and the likelihood could be expressed as

$$L(\beta,D,\phi) = \prod_{i=1}^{N} f_i(\mathbf{y_i}|\beta,\mathbf{D},\phi).$$

However, this likelihood does not always have a closed form solution. As in the case of binary responses, approximations were required. The method employed in this report was the Penalized Quasi-Likelihood (PQL) approach, wherein the mean function, μ_{ij}, was approximated through a linear Taylor expansion using the current estimates, $\hat{\beta}$ and $\hat{b}_i$, yielding a pseudo-response,

$$\mathbf{Y_i^*} \equiv \mathbf{X_i}\hat{\beta} + \mathbf{Z_i}\hat{\mathbf{b}}_\mathbf{i} + \hat{\mathbf{V}}_\mathbf{i}^{-1}(\mathbf{Y_i} - \hat{\mu}_\mathbf{i}).$$

Model fitting was done by iteratively updating the pseudo-responses and fitting the following linear mixed model to them until convergence was reached

$$\mathbf{Y}_\mathbf{i}^* \approx \mathbf{X}_\mathbf{i}\beta + \mathbf{Z}_\mathbf{i}\mathbf{b}_\mathbf{i} + \varepsilon_\mathbf{i}^*.$$

2.3 Experimental Settings, Corresponding Models and Data Simulation

It should be recalled that the experimental design factors considered in this report were location, experimental unit, type of trial and type of outcome, of two types each, generating 16 experimental settings (Table 1). Data simulation and corresponding linear mixed models were unique for each of these settings, thus 16 SAS macros were constructed in order to make power calculation possible for all of them.

Moreover, within each experimental setting, three blocking designs could be employed in conducting the trial namely, CRD, RCBD and GRBD. Therefore, each of the macro was made such that it could further address power calculations when the trials are planned with these blocking designs in mind. The main difference with CRD and the blocked designs (RCBD and GRBD) was the absence of blocking. Thus, for all settings, data simulation included generation of blocks, but it should be noted that for CRD settings, this step was not done. It should also be emphasized that RCBD differed from GRBD such that the blocks in RCBD designs would only contain one experimental unit per treatment within the block.

Setting		Site	EU	Type of trial	Type of data	Random Effects
A	1	Single	Animal	Superiority	Continuous	Block
	2				Binomial	
	3			Non-inferiority	Continuous	
	4				Binomial	
B	5		Pen	Superiority	Continuous	Block, Pen(Trt*Block)
	6				Binomial	
	7			Non-inferiority	Continuous	
	8				Binomial	
C	9	Multi	Animal	Superiority	Continuous	Site, Block(Site)
	10				Binomial	Site*Trt
	11			Non-inferiority	Continuous	
	12				Binomial	
D	13		Pen	Superiority	Continuous	Site, Block(Site),
	14				Binomial	Pen(Trt*Block*Site),
	15			Non-inferiority	Continuous	Site*Trt
	16				Binomial	

****CRD - no block random effect.*

Table 1: *List of Experimental Settings and Identifying Design Factors*

2.3.1 Continuous Response

The general model for the continuous response is given by the following:

$$Y_{ijklm} = \mu + \tau_i + \gamma_j + \tau\gamma_{ij} + \beta_{k(j)} + \pi_{l(ijk)} + \varepsilon_{ijklm} \qquad (3)$$

where,

Y_{ijklm} =observation for the m^{th} animal in the l^{th} pen within the k^{th} block, i^{th} treatment and j^{th} center

μ =overall constant

τ_i =fixed effect of the i^{th} treatment

γ_j =random effect of the j^{th} center

$\tau\gamma_{ij}$ =random interaction effect of the i^{th} treatment and j^{th} center

$\beta_{k(j)}$ =random effect of the k^{th} block within the j^{th} center

$\pi_{l(ijk)}$ =random effect of the l^{th} pen within the k^{th} block, i^{th} treatment and j^{th} center

ε_{ijklm} =residual

i =1,2

j =1,.., number of centers

k =1,.., number of blocks within treatment and center

l =1,.., number of pens within block, treatment and center

m =1,.., number of animals within a pen

The model above described the most complex setting wherein the trial would have a multi-center location, pen as the experimental unit, within which, several animals could be housed, and a blocking design would be employed, allowing several pens within it. This was expressed by the inclusion of the center, center by treatment interaction, pen and block random effects. The nested design of the experiment was appropriately described by the subscripts in the notation. Simpler models for other experimental settings could be expressed as a simplification of Equation 3. It should be noted that the center by treatment interaction was included in the model to take into account the possible differences in the treatment effect between centers[9], and is also advised by regulatory bodies for veterinary clinical trial conduct[5].

Only two treatments were compared as the primary goal was to demonstrate either superiority or non-inferiority of an investigational drug compared to a reference treatment. All designs were balanced such that the number of experimental units per block, per treatment, and per center were equal. And the number of blocks per treatment and per center were also equal in all designs. For all data simulations, the macro would require a value for μ, which was the mean for the response in the reference group, and $\mu + delta$, as mean for the response in the treatment group. It should be noted that delta was the expected treatment difference between the two treatment arms desired to be detected by the trial.

Data simulation was done according to the structure of the design implied by the experimental setting and the corresponding model. Centers were generated and the corresponding center random effects were drawn from $N(0,\sigma_c^2)$. Within each center, blocks were generated and corresponding block random effects were drawn from $N(0,\sigma_b^2)$. Within each block, treatments were assigned to pens with their corresponding mean response and their center by treatment interaction drawn from $N(0,\sigma_{ct}^2)$. The pens generated for every treatment were assigned a random pen effect drawn from $N(0,\sigma_p^2)$. Finally, within each pen, animals were generated with the corresponding error terms drawn from $N(0,\sigma^2)$. The final response consisted of the sum of the mean response, the random effects of center, center by treatment interaction, block and pen and lastly, the residual. Variance parameters required were variance between centers, variance of the center by treatment interaction, variance between blocks, variance between pens and variance of residuals.

2.3.2 Binary Response

It should be recalled that the form of the generalized linear mixed model employed in the binary data case was given by Equation (2) in Section 2.2.2. For the most complex setting, wherein the trials were conducted in multiple centers, with GRBD as the blocking design and pen was the experimental unit, the model was given by:

$$Y_{ijklm}|\mathbf{b} \sim Bernoulli(\pi_{ijklm})$$
$$logit(\pi_{ijklm}) = \mu + \tau_i + \gamma_j + \tau\gamma_{ij} + \beta_{k(j)} + \rho_{l(ijk)},$$

where,

Y_{ijklm} =observation for the m^{th} animal in the l^{th} pen within the k^{th} block, i^{th} treatment and j^{th} center

π_{ijklm} =probability of success for the m^{th} animal in the l^{th} pen within the k^{th} block, i^{th} treatment and j^{th} center

μ =overall constant

τ_i =fixed effect of the i^{th} treatment

γ_j =random effect of the j^{th} center

$\tau\gamma_{ij}$ =random interaction effect of the i^{th} treatment and j^{th} center

$\beta_{k(j)}$ =random effect of the k^{th} block within the j^{th} center

$\rho_{l(ijk)}$ =random effect of the l^{th} pen within the k^{th} block, i^{th} treatment and j^{th} center

i =1,2

j =1,.., number of centers

k =1,.., number of blocks within treatment and center

l =1,.., number of pens within block, treatment and center

m =1,.., number of animals within pen

The drawing of the random effects was similar with what was described previously for the continuous case. The center, center by treatment interaction, block and pen random effects were drawn from normal distributions with mean zero and the pre-specified variances. However, the random effects were added to the logit scale, thus, the generation of the individual animal response was quite different for the binary response, and would be illustrated below.

For clinical trials, though binary response data were in terms of Yes/No or 0/1, the results of the analysis would be usually presented in proportions or rate such as mortality rate or cure rate. Thus, unlike in the continuous case where μ denoted the mean response in the reference group, in the case of the binary data,

$$\mu_{ref} = logit(\pi_{ref}),$$

wherein π_{ref} is the probability of success in the reference group and μ_{ref} is the equivalent value in the logit scale . In the treatment group, the probability of success could be denoted by π_{trt}, and could actually be expressed as $\pi_{ref} + delta$, where delta is the expected difference in the probability of successes in the two groups. Thus, π_{trt} could be expressed in the logit scale as

$$\mu_{trt} = logit(\pi_{trt}) = logit(\pi_{ref} + delta).$$

This μ_{ref} and μ_{trt} were the mean responses in the logit scale to which the random effects were added. The usual data simulation scheme was employed for each setting wherein the necessary random effects were drawn and added to these mean responses depending on treatment. After this sum was generated, it was transformed back to its probability scale by the expit function as shown below for an animal in the reference group,

$$\pi_{ref} = \frac{exp(\mu_{ref} + z_{center} + z_{center*trt} + z_{block} + z_{pen})}{1 + exp(\mu_{ref} + z_{center} + z_{center*trt} + z_{block} + z_{pen})}.$$

where,

$$\begin{aligned} z_{center} &\sim N(0, \sigma_c^2) \\ z_{center*trt} &\sim N(0, \sigma_{ct}^2) \\ z_{block} &\sim N(0, \sigma_b^2) \\ z_{pen} &\sim N(0, \sigma_p^2) \end{aligned}$$

This probability then was used to draw a response, Y, from a Bernoulli distribution with a parameter π_{ref}. The procedure for the treatment group was identical. It should be noted that for binary responses, no random errors from a normal distribution were drawn. The last source of variation for the animal response in the simulations was the generation of the response from a Bernoulli distibution with a probability parameter which was determined from sum of the mean response in the logit scale and the included random terms.

2.4 Power Calculation

Power is the probability of rejecting the null hypothesis when the alternative hypothesis is true. Thus, in the settings considered, it is the measure of the ability of the study design to detect a clinically meaningful treatment effect, which would warrant the approval of an experimental drug. This power is dependent on the hypothesis to be tested, study design, sampling design and the statistical method to be employed in the analysis [10]. For some statistical models and tests, definite formulations or approximations of the distribution of the test statistic exist such that power analysis could be done by plugging parameter values to the mathematical formula, and power is readily calculated. For fixed effects models, for instance, power could be calculated exactly through noncentral F and noncentral t-distributions for many special cases such as t -tests and ANOVA [4]. For the case of mixed models however, the distribution of the test statistic is usually only known under the null hypothesis [24].

It is described in detail in Verbeke (2000) and supported by simulations in Helms (1992) that the distribution of the F-statistic for the test of the linear combination of fixed effects could be approximated by an F-distribution with a non-centrality parameter. One then would only need to sample an ideal data set once, fit the mixed model of choice, generate the non-centrality parameter and degrees of freedom, and determine the F-quantile, which would give the probability of correctly rejecting the null hypothesis under the alternative hypothesis [24]. However, simulations as a means of power calculation is always a valid approach and may prove to be more accurate than approximations when repeated for a large number of times[4]. Thus, in this report, this latter option was employed. In addition, non-inferiority testing involve construction of confidence intervals and comparison of confidence limits with clinically acceptable margin which was not straightforward in the previously method mentioned.

Power calculations were conducted by simulating 1000 data sets structured according to designs described in Section 2.3. The sources of variation immediately followed from the design, and thus, variances were pre-specified before the simulations. Appropriate mixed models were fitted to the data sets, and superiority or non-inferiority testing was done depending on the objective. The details of the tests would be discussed below, but the main goal was calculating the percentage of correctly rejecting the null hypothesis for all 1000 data sets simulated, giving the approximate power of the specific experimental design considered.

2.4.1 Superiority Tests

When the aim of the trial is to exhibit superiority of an experimental drug over a standard drug, one is interested in testing whether the clinically relevant treatment difference, delta, is significantly different from 0. Thus, for a two-sided superiority trial, the hypotheses tested were

$$H_o : \mu_{trt} - \mu_{ref} = \Delta = 0$$
$$H_a : \mu_{trt} - \mu_{ref} = \Delta \neq 0$$

a. Continuous Outcome

Using the SAS contrast statements, this testing was done by employing the approximate F-tests for linear combination of fixed effects proposed in Duchateau (1997), wherein to test the general linear hypothesis of the of the form,

$$H_o : \mathbf{C}'\beta = \mathbf{0}$$
$$H_a : \mathbf{C}'\beta \neq \mathbf{0}$$

the distribution of the test statistic,

$$F = \frac{(C'\hat{\beta})'(C'(X'\hat{V}^{-1}X)^{-1}C)^{-1}(C'\hat{\beta})}{rank(C)},$$

is approximated under null by an F-distribution with numerator degrees of freedom equal to rank(C) and denominator degrees of freedom to be approximated from the data.

b. Binary Outcome

The same procedure for testing the treatment effects was done for the binary outcome using the approximate F-test for the fixed effects described above for the continuous case. Molenberghs

(2005) noted that since most generalized mixed models parameters are estimated by fitting linear mixed models to pseudo-data, which was the case of the simulations done, approximate F and t-tests for fixed effects directly followed from the linear mixed model framework[19].

In the simulations, significance of treatment difference, and consequently, superiority of treatment were exhibited when p-values for the mentioned two-sided tests were less than or equal to the specified α level. In this report, α was set to 0.05. Out of the 1000 data sets, 1000 corresponding analysis were done and the percentage of significant results gave the power estimate for the experimental design considered.

2.4.2 Non-Inferiority Tests

Non-inferiority trials are conducted to exhibit that the experimental drug has a comparable efficacy to that of a standard treatment. This is established by choosing Δ_{NI}, which is the clinically acceptable margin of difference, and testing the following hypothesis

$$H_o : \mu_{trt} - \mu_{ref} = \Delta > \Delta_{NI}$$
$$H_a : \mu_{trt} - \mu_{ref} = \Delta < \Delta_{NI}.$$

Nowadays, the standard non-inferiority tests are performed at a one-sided 0.025 level, and results are reported through confidence intervals [17]. In this report, $-\Delta_{NI}$ determined the lower bound for the acceptable margin of treatment difference. When the lower limit of the one-sided $(1-\alpha)$ confidence interval for the treatment difference was above it, the treatment was demonstrated to be non-inferior.

a. Continuous Outcome
Since it was possible to conduct approximate t-tests for fixed effects in linear mixed models [24], generation of the two-sided $(1-\alpha)$ confidence interval was straightforward for continuous outcomes in SAS by using the *Estimate* option in Proc Mixed. Thus, for a non-inferiority test with $\alpha = .025$, where the lower limit of the one-sided 97.5% CI for treatment difference should be determined, the equivalent lower limit of a two-sided 95% CI was generated from SAS. Non-inferiority of treatment was demonstrated when this mentioned limit was found to be greater than $-\Delta_{NI}$.

b. Binary Outcome

i. Confidence Interval for Difference of Proportion (Independence Assumption)
The first approach on non-inferiority testing proposed in this report was by constructing confidence intervals around the difference of proportions, $\pi_{trt} - \pi_{ref}$, with standard errors computed assuming independence of the two proportions. Analysis were constrained for the fixed effects by setting random effects equal to zero and from Equation 2, the proportion immediately followed as

$$\pi = \frac{\mathbf{exp(X}\beta)}{\mathbf{1+exp(X}\beta)}.$$

The standard error assuming independence was given by

$$\hat{\sigma}(\pi_{trt} - \pi_{ref}) = \sqrt{\frac{\pi_{trt}(1-\pi_{trt})}{n_{trt}} + \frac{\pi_{ref}(1-\pi_{ref})}{n_{ref}}}.$$

Constructing the Wald confidence interval by substituting π_i with $\hat{\pi}_i$ would give

$$\hat{\pi}_{trt} - \hat{\pi}_{ref} \pm Z_{\frac{\alpha}{2}} \hat{\sigma}(\hat{\pi_{trt}} - \hat{\pi_{ref}})$$

However, Agresti noted that this interval performs poorly for small n[1]. In this report then, the t-statistic was used to build the CI as this would be a more conservative approach than employing the Z-statistic for small sample sizes, and would of course generate approximately the same CI when the sample size is large. The degrees of freedom used was equal to the denominator degrees of freedom approximated for the t-tests on treatment fixed effect with the motivation that the same sources of variation specific for the design should be accounted for when testing the difference of proportions.

The alternative C.I considered in this report then was

$$\hat{\pi}_{trt} - \hat{\pi}_{ref} \pm t_{\frac{\alpha}{2}}(v) \hat{\sigma}(\hat{\pi_{trt}} - \hat{\pi_{ref}})$$

ii. Confidence Interval for Difference of Proportion (Delta Method)
The method above assumed independence of the proportions from the treatment and the reference groups. This, of course, was a very strong assumption. Alternatively, another way of approximating variances was through the delta method. Agresti (2006) elaborately discussed how it could be done to several logistic parameters. In a nutshell, when T_n is asymptotically normal, wherein

$$\sqrt{n}(T_n - \theta) \overset{d}{\rightarrow} N(0, \sigma^2),$$

then an estimator $g(T_n)$ which is a function of T_n is also asymptotically normal such that

$$\sqrt{n}[g(T_n) - g(\theta)] \overset{d}{\rightarrow} N(0, [g'(\theta)]^2 \sigma^2).$$

Molenberghs (2005) noted that for fixed effects in GLMMs, asymptotic normality follows from the central limit theorem on $\hat{\beta}$. Since π_{trt} and π_{ref} are just functions of parameters in $\hat{\beta}$, the delta method on deriving standard errors offered a reasonable alternative so that the independence assumption in the previously mentioned method would not be necessary anymore. Within this framework, the standard error of $(\pi_{trt} - \pi_{ref})$ was derived. The derivation proceeded by letting

$$\pi_{trt} = \frac{exp(\alpha + \beta)}{1 + exp(\alpha + \beta)} \quad \text{and} \quad \pi_{ref} = \frac{exp(\alpha)}{1 + exp(\alpha)} \tag{4}$$

Through delta method the variance of $\pi_{trt} - \pi_{ref}$ was approximated below by

$$\begin{aligned} V_d(\pi_{trt} - \pi_{ref}) = & \left[\frac{\partial(\pi_{trt} - \pi_{ref})}{\partial \alpha}\right]^2 V(\alpha) + \left[\frac{\partial(\pi_{trt} - \pi_{ref})}{\partial \beta}\right]^2 V(\beta) \\ & + 2\frac{\partial(\pi_{trt} - \pi_{ref})}{\partial \alpha}\frac{\partial(\pi_{trt} - \pi_{ref})}{\partial \beta} Cov(\alpha, \beta). \end{aligned}$$

Detailed derivations were included in the Appendix, and it could be shown that,

$$\begin{aligned} \frac{\partial(\pi_{trt} - \pi_{ref})}{\partial \alpha} = & \pi_{trt}(1 - \pi_{trt}) - \pi_{ref}(1 - \pi_{ref}) \\ \frac{\partial(\pi_{trt} - \pi_{ref})}{\partial \beta} = & \pi_{trt}(1 - \pi_{trt}), \end{aligned}$$

and the approximated variance for the difference of proportions would be given by

$$\begin{aligned} V_d(\pi_{trt} - \pi_{ref}) = & \pi_{trt}^2(1 - \pi_{trt})^2 V(\alpha + \beta) + \pi_{ref}^2(1 - \pi_{ref})^2 V(\alpha) \\ & - 2\pi_{trt}(1 - \pi_{trt}) * \pi_{ref}(1 - \pi_{ref})\Big[V(\alpha) + Cov(\alpha, \beta)\Big]. \end{aligned}$$

However, also through delta method and were shown in the Appendix, the individual variances of π_{trt} and π_{ref} could be approximated by

$$\begin{aligned} V_d(\pi_{trt}) = & \pi_{trt}^2(1 - \pi_{trt})^2 V(\alpha + \beta) \\ V_d(\pi_{ref}) = & \pi_{ref}^2(1 - \pi_{ref})^2 V(\alpha), \end{aligned}$$

such that the variance of the difference of proportions could be further expressed in terms of the variances of the individual proportions,

$$V_d(\pi_{trt} - \pi_{ref}) = V_d(\pi_{trt}) + V_d(\pi_{ref}) - 2\pi_{trt}(1 - \pi_{trt}) * \pi_{ref}(1 - \pi_{ref})\Big[V(\alpha) + Cov(\alpha, \beta)\Big].$$

SAS could generate estimates for all the terms in the equation above. Thus, this approximated standard error could be calculated when fitting GLMMs giving the following $(1 - \alpha)$ two-sided confidence interval,

$$\hat{\pi}_{trt} - \hat{\pi}_{ref} \pm Z_{\frac{\alpha}{2}}\sqrt{\hat{V}_d(\hat{\pi_{trt}} - \hat{\pi_{ref}})}.$$

Alternatively, to take into account the design and the presence of random effects in the model, the CI constructed using the t-distribution with degrees of freedom equal to the one approximated for the test of the treatment fixed effect was used with the same motivation above, and was given by the following,

$$\hat{\pi}_{trt} - \hat{\pi}_{ref} \pm t_{\frac{\alpha}{2}}(v)\sqrt{\hat{V}_d(\hat{\pi_{trt}} - \hat{\pi_{ref}})}.$$

iii. Confidence Interval for Odds Ratio

The last approach investigated in this report to test non-inferiority for binary outcome was testing through CIs of Odds ratios. It should be recalled when proportions could be expressed as in Equation 4, the odds ratio was given by

$$OR = exp(\beta).$$

Agresti noted that log(OR) is approximately normal and its $(1-\alpha)$ two-sided confidence interval could be approximated by

$$log(\hat{OR}) \pm z_{\frac{\alpha}{2}}\sqrt{\hat{V}(log(\hat{OR}))},$$

and the $(1-\alpha)$ two-sided confidence interval for the OR is calculated by exponentiating the limits of this interval. In the case of GLMMs fitted in this report, $log(OR)$, which was equal to β, have the following $(1-\alpha)$ two-sided confidence interval, wherein the distribution used was the t-distribution with the approximated degrees of freedom for this fixed effect,

$$\hat{\beta} \pm t_{\frac{\alpha}{2}}(v)\sqrt{\hat{V}(\hat{\beta})},$$

The $(1-\alpha)$ two-sided confidence interval for OR then was derived by exponentiating the corresponding limits.

Furthermore, since the acceptable margin of difference, Δ_{NI}, was expressed in difference of proportions, there was a need to express this limit in terms of odds ratio, which was done in the following manner,

$$\Delta_{Odds-NI} = \frac{\frac{\pi_{ref}-\Delta_{Ni}}{1-(\pi_{ref}-\Delta_{Ni})}}{\frac{\pi_{ref}}{1-\pi ref}}$$

Testing was done by comparing the lower limit of the $(1-\alpha)$ two-sided confidence interval generated for the odds ratio to $\Delta_{Odds-NI}$. When the lower limit was greater than $\Delta_{Odds-NI}$, the treatment showed non-inferiority.

All confidence interval construction described were for a two-sided $(1-\alpha)$ CI, however, it should be noted that only the lower limit of these CIs were compared to the acceptable clinical margin of difference such that the actual inferiority tests conducted were one-sided with significance level equal to $\frac{\alpha}{2}$. For instance, for non-inferiority tests with 97.5% level of confidence, two-sided 95% CI lower limits were compared to Δ_{NI}. But this lower limit was still equal to the lower limit of a one-sided 97.5% C.I, thus, the testing was still done in accordance to the non-inferiority test desired. Similar with the procedure employed in the superiority tests, power was estimated by calculating the percentage of significantly non-inferior results over the total number of simulated data sets.

3 Results

3.1 Superiority Tests

3.1.1 Single Center - Animal as EU - Continuous Response

Through the simulations, power was calculated for the different experimental designs. The parameters used for the first setting, where block was the only random effect, were tabulated in Table 2. It should also be noted that for the data in the CRD setup, no block variability was included during data generation, and thus, they were less variable. The simulations were done in such a way that the absolute treatment difference (Δ) and the variance of the residual (σ^2_{res}) were the same for all blocking designs.

For a standardized delta (Δ_{std}) of 50%, more or less 30 animals for each treatment were needed to achieve an acceptable power of 80% (Table 3). It could be seen that across the three blocking designs, power was still comparable even though data from the blocked designs had a larger total variance. It is thus shown how blocking helped to generate efficient estimates for the treatment difference in the two scenarios considered, wherein the between block variance constituted 60% of the total variance. Table 3 also showed that for a smaller treatment difference desired to be detected, $\Delta_{std} = 25\%$ for instance, the lower was the power for the design. More subjects then should be included in the study to achieve an acceptable power.

Parameter	Scenario 1				Scenario 2 - Δ decreased			
	RCBD/ GRBD		CRD		RCBD/GRBD		CRD	
σ^2_{block}	0.15	60%			0.15	60%		
σ^2_{res}	0.10	40%	0.10	100%	0.10	40%	0.10	100%
σ^2_{total}	0.25	100%	0.10	100%	0.25	100%	0.10	100%
σ_{total}	0.50		0.32		0.50		0.32	
Δ	0.25		0.25		0.125		0.125	
$\Delta_{std} = \frac{\Delta}{\sigma_{total}}$	0.50		0.79		0.25		0.40	

Table 2: *Pre-specified Parameters in Setting A: Single Center, Animal as EU, Continuous Outcome*

Furthermore, varying the values for intra-block correlation, given by,

$$\rho_b = \frac{\sigma^2_{block}}{\sigma^2_{total}}$$

was done to investigate on its consequences on power. Using the parameters in Scenario 1 from Table 2 and a GRBD design having five blocks per treatment, it could be seen from the simulation results in Figure 1 that the greater was the intrablock correlation, wherein subjects within a block were more homogeneous, the more powerful was the design.

Design	Blocks	Animal/ Block/ Treatment	Animal/ Treatment	Convergence (%)	Power ($\Delta_{std} = 50\%$)	Power ($\Delta_{std} = 25\%$)
CRD	-	12	12	-	44.5	15.0
	-	20	20	-	69.2	23.1
	-	30	30	-	83.8	31.6
	-	40	40	-	92.8	42.7
GRBD	3	4	12	100.0	45.3	14.1
	5	4	20	100.0	66.7	20.7
	6	5	30	100.0	86.5	31.1
	10	4	40	100.0	92.9	39.5
RCBD	12	1	12	100.0	41.3	14.2
	20	1	20	100.0	67.3	22.8
	30	1	30	100.0	82.7	32.0
	40	1	40	100.0	93.7	41.0

****For CRD: $\Delta_{std} = 80\%, 40\%$, respectively. Convergence status is NA: there are no random effects included.*

Table 3: *Power for Setting A: Single Center, Animal as EU, Continuous Outcome, Superiority*

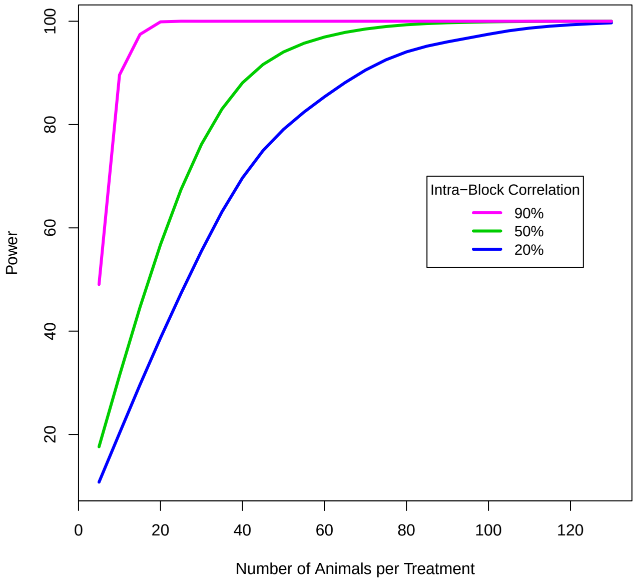

Figure 1: *Power for Setting A: Single Center, Animal as EU, Continuous Outcome, Superiority with varying Intrablock Correlation (GRBD with 5 blocks)*

3.1.2 Single Center - Pen as EU - Continuous Response

For the third setting, the main design feature was having pen as the experimental unit. Two scenarios were looked at to compare results if σ^2_{pen} was increased, and consequently,

$$\rho_p = \frac{\sigma^2_p}{\sigma^2_{total}},$$

such that animals within a cluster were more correlated (Table 6).

Parameter	Scenario 1				Scenario 2 - σ^2_{pen} increased			
2-9	RCBD/ GRBD		CRD		RCBD/GRBD		CRD	
σ^2_{pen}	0.15	38%	0.15	60%	0.24	60%	0.24	96%
σ^2_{block}	0.15	38%		0%	0.15	38%		0%
σ^2_{res}	0.10	25%	0.10	40%	0.01	3%	0.01	4%
σ^2_{total}	0.40	100%	0.25	100%	0.40	100%	0.25	100%
σ_{total}	0.63		0.50		0.63		0.50	
Δ_{abs}	0.50		0.50		0.50		0.50	
$\Delta_{std} = \frac{\Delta}{\sigma_{total}}$	0.79		1.00		0.79		1.00	

Table 4: *Pre-specified Parameters in Setting B: Single Center, Pen as EU, Continuous Outcome*

							Power ($\Delta_S = 80\%$)	
Design	Blocks	Pen/ Block/ Treatment	Animal/ Pen/ Block/ Treatment	Pen/ Treatment	Animal/ Treatment	Convergence (%)	σ_{pen}: 38%	σ_{Pen}:60%
CRD	-	2	8	2	16	100.0	9.8	8.6
	-	2	16	2	32	100.0	11.6	9.2
	-	2	40	2	80	100.0	12.0	8.4
	-	8	2	8	16	100.0	54.1	46.6
	-	16	2	16	32	100.0	85.8	77.5
	-	40	2	40	80	100.0	99.9	99.5
GRBD	2	4	2	8	16	100.0	52.0	45.3
	2	8	2	16	32	100.0	85.3	79.3
	2	20	2	40	80	100.0	99.9	99.3
RCBD	8	1	2	8	16	100.0	45.5	38.8
	16	1	2	16	32	100.0	81.7	75.7
	40	1	2	40	80	100.0	99.8	99.1

****For CRD: σ^2_{pen}: 60%, 96%, respectively.*

Table 5: *Power for Setting B: Single Center, Pen as EU, Continuous Outcome, Superiority*

Unlike in the previous settings where power was increased when between block variability was higher, in this setting, the more variation was present between the pens, implying higher intracluster correlation, the lesser was the power of the design (Table 5). In addition, focusing on the settings for the CRD design, where simulations were done to compare the consequences on power if the number of animals per pen was increased and the number of pens was fixed to 2, it could be seen that power did not improve substantially even if the number of animals was increased tremendously such that it was doubled or five-fold. However, a drastic increase in power was generated for simulated trials with the same number of animals but with more pens.

Further comparison was done to observe power when the number of pens were fixed and the number of animals per pen were increased. It could be seen from Figure 2 that even when the number of animals per pen were increased in extremely large values, power seemed to stabilize and reach a certain ceiling. Specifically for trials with 10 pens per block with parameters equal to that of Scenario 1 (Table 4) with pen variability comprising 38% of the total variability, power did not reach the acceptable level of 80% even when there were already 30 animals per pen included or an equivalent of 600 animals for a trial employing GRBD with 2 blocks design.

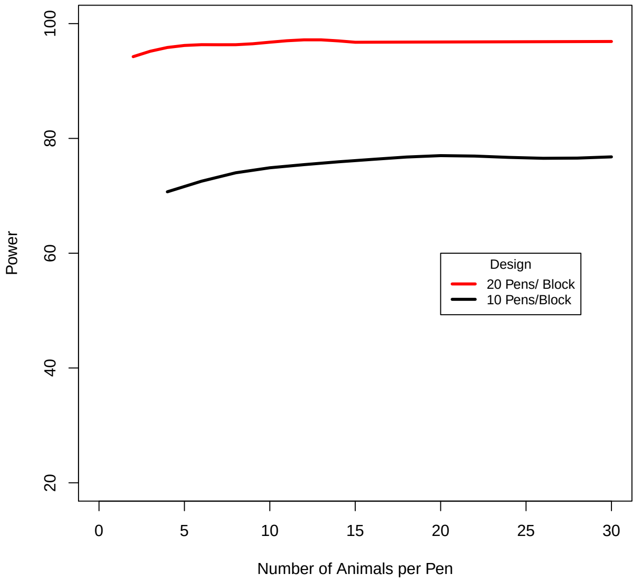

Figure 2: *Power for Setting B: Single Center, Pen as EU, Continuous Outcome, Superiority with varying Number of Animals per Pen (GRBD with 2 Blocks)*

3.1.3 Multi-Center - Animal as EU - Continuous Response

The second setting considered for the simulations have parameters used for the simulations listed in Table 4. For Scenario 2, σ^2_{center} was increased, and for Scenario 3, $\sigma^2_{center*trt}$ was increased to observe their influences on power of the design.

Parameter	Scenario 1				Scenario 2 - σ^2_{center} increased				Scenario 3 - $\sigma^2_{center*trt}$ increased			
	RCBD/ GRBD		CRD		RCBD/ GRBD		CRD		RCBD/ GRBD		CRD	
σ^2_{center}	0.04	13%	0.04	27%	0.10	33%	0.10	67%	0.04	13%	0.04	27%
$\sigma^2_{center*trt}$	0.01	3%	0.01	7%	0.01	3%	0.01	7%	0.06	20%	0.06	40%
σ^2_{block}	0.15	50%			0.15	50%			0.15	50%		
σ^2_{res}	0.10	33%	0.10	67%	0.04	13%	0.04	27%	0.05	17%	0.05	33%
σ^2_{total}	0.30	100%	0.15	100%	0.30	100%	0.15	100%	0.30	100%	0.15	100%
σ_{total}	0.55		0.39		0.55		0.39		0.55		0.39	
Δ_{abs}	0.275		0.275		0.275		0.275		0.275		0.275	
$\Delta_{std} = \frac{\Delta}{\sigma_{total}}$	0.50		0.71		0.50		0.71		0.50		0.71	

Table 6: *Pre-specified Parameters in Setting C: Multi-Center, Animal as EU, Continuous Outcome, Superiority*

As with blocking, taking into account the effect of center proved to be beneficial for the power of a design. It could be seen in Table 7 that when there was a considerable center variability, power was increased as more variation attributed to center was taken into account, and the corresponding significance tests had a lower residual variability.

Since the models took into account the center by treatment effect, the tests for significance were dependent on $\sigma^2_{center*trt}$, and the corresponding degrees of freedom would follow on the number of centers included in the study. More centers would give more degrees of freedom to the significance tests, thus giving them more power. This is shown in Figure 3, wherein given the same number of animals, simulated trials which had 10 centers, easily achieved higher power compared to those with 5 centers only.

The effect of increasing the center by treatment interaction variability on power was also investigated. In Table 7, it could be seen that in Scenario 3, when the center by treatment interaction variability was increased, implying a greater degree of differences in the effect of treatment across centers, power was reduced. Again, since $\sigma^2_{center*trt}$ was not separated from σ^2 when testing for the significance of treatment effect, it added to the noise that blurred the estimation of the effect, thus, reducing the power.

Design	Centers	Block/ Center/ Treatment	Animals/ Block/ Treatment	Animals/ Center/ Treatment	Animals/ Treatment	Convergence (%)	Power ($\Delta_{std} = 50\%$)		
							Scenario 1	Scenario 2 σ^2_{center} increased	Scenario 3 $\sigma^2_{center*trt}$ increased
CRD	3	-	4	4	12	100.0	2.4	13.8	7.6
	5	-	4	4	20	100.0	34.3	61.2	22.0
	10	-	3	3	30	100.0	74.5	95.3	50.9
	10	-	4	4	40	100.0	81.7	97.0	51.8
	20	-	4	4	80	100.0	99.7	100.0	86.1
GRBD	3	2	2	4	12	100.0	4.0	14.7	7.9
	5	2	2	4	20	100.0	38.3	73.0	33.3
	8	2	2	4	32	100.0	75.4	96.6	63.3
	10	2	2	4	40	100.0	87.4	99.0	75.4
	20	2	2	4	80	100.0	99.8	100.0	97.6
RCBD	3	4	1	4	12	100.0	4.4	15.9	3.6
	5	4	1	4	20	100.0	41.4	78.7	41.0
	10	3	1	3	30	100.0	78.4	99.0	78.5
	10	4	1	4	40	100.0	89.7	99.8	88.5
	20	4	1	4	80	100.0	99.8	100.0	99.8

****For exact proportion of variances, please refer to Table 6*

Table 7: *Power for Setting C: Multi-Center, Animal as EU, Continuous Outcome, Superiority*

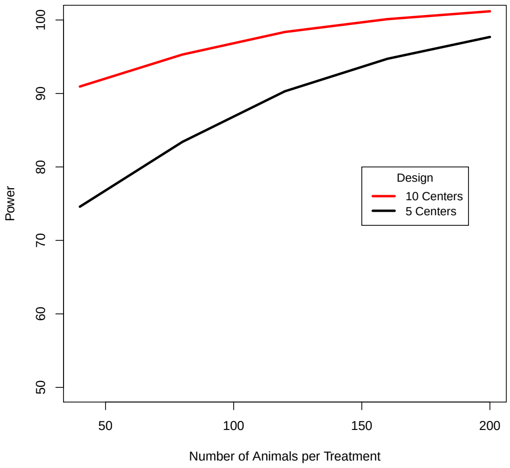

Figure 3: *Power for Setting C: Multi-Center, Animal as EU, Continuous Outcome, Superiority with varying Number of Centers (GRBD with 2 blocks per Center)*

3.1.4 Multi-Center Pen as EU - Continuous Response

The same findings regarding the effects of varying the between block, center and pen variabilities were observed for the last and the most complex setting which were trials having multi-center location and pen as the experimental unit. Detailed results could be found in Table A2 in the Appendix. Results for superiority testing in the binary settings were also tabulated in the mentioned section (Tables A3-A6).

3.2 Non-inferiority Tests

For non-inferiority tests of continuous outcomes, it was already mentioned in the methodology that the method was straightforward as SAS generated confidence intervals for treatment differences using the approximate t-distributions. The results then were similar to what was observed with the superiority tests described above. Results exhibited below then were from non-inferiority testing of binary responses where three methods of constructing confidence intervals were employed.

3.2.1 Single Center Animal as EU - Binary Response

For the most simple setting wherein simulated trials were single-centered and animal was the experimental unit, it is shown in Table 8 that non-inferiority tests using CI for $\pi_{trt} - \pi_{ref}$ (Independence) and CI for $\pi_{trt} - \pi_{ref}$ (Delta Method) generated very similar results, with the latter consistently having the higher value. CI for odds ratio, on the other hand, proved to be the most conservative of all methods giving the lowest powers in all scenarios simulated as this CI is usually wider because of the skewness resulting from the transformation from the logit scale [11]. These findings were also observed in Settings B and C shown in Tables A7-A8 in the Appendix.

					Power		
Design	Block	Animals/ Block/ Treatment	Animals/ Treatment	Convergence (%)	$\pi_{trt} - \pi_{ref}$ (Indep)	$\pi_{trt} - \pi_{ref}$ (Delta)	Odds Ratio
CRD	-	12	12	100.0	19.7	19.7	10.4
	-	20	20	100.0	27.3	27.3	18.3
	-	40	40	100.0	49.4	49.4	36.8
	-	80	80	100.0	77.4	77.4	61.6
GRBD	6	2	12	96.0	12.5	15.0	9.9
	10	2	20	100.0	22.5	27.0	16.7
	20	2	40	100.0	45.1	48.1	37.2
	40	2	80	98.2	77.5	79.6	64.1
RCBD	12	1	12	96.3	14.6	16.2	10.8
	20	1	20	99.6	22.5	24.5	17.9
	40	1	40	100.0	47.9	49.7	38.3
	80	1	80	100.0	79.2	80.3	64.8

$\pi_{trt} = 0.65$, $\pi_{ref} = 0.75$, $\Delta_{NI} = 0.30$, $\sigma^2_{block} = 0.15$

Table 8: *Power for Setting A: Single Center, Animal as EU, Binary Outcome, Non-Inferiority*

3.2.2 Multi-Center Pen as EU - Binary Response

For the most complex setting, however, where simulated trials were done in multi-centers with pen as the experimental unit, results in Table 9 showed that non-inferiority tests using the CI for $\pi_{trt} - \pi_{ref}$ (Independence)was the most conservative, generating most of the the lowest powers across the different settings. The assumption of independence then became too stringent for this design. The highest powers were still derived for tests which used the CI for $\pi_{trt} - \pi_{ref}$ (Delta Method). And CI for OR had power values which were usually within the generated estimates from the two previously mentioned methods.

								Power		
Design	Centers	Block/ Center	Pens/ Block/ Trt	Animals/ Pen	Animals/ Trt	Pens/ Trt	Convergence (%)	$\pi_{trt} - \pi_{ref}$ (Indep)	$\pi_{trt} - \pi_{ref}$ (Delta)	Odds Ratio
CRD	4	-	4	2	32	16	99.8	4.5	6.3	3.4
	4	-	10	2	80	40	96.1	14.7	26.4	15.6
	4	-	20	2	160	80	91.3	24.2	59.7	43.2
	8	-	10	2	160	80	92.4	70.3	88.2	74.7
GRBD	4	2	2	2	32	16	98.9	4.2	6.5	2.2
	4	2	5	2	80	40	96.6	9.9	25.5	12.7
	4	2	10	2	160	80	85.2	19.1	62.7	44.8
	8	2	5	2	160	80	93.0	65.1	88.0	76.0
RCBD	4	4	1	2	32	16	99.0	4.0	6.6	2.9
	4	10	1	2	80	40	93.1	12.4	26.0	16.6
	4	20	1	2	160	80	83.7	27.5	61.8	45.5
	8	10	1	2	160	80	89.8	72.4	90.5	79.3

:$\pi_{trt} = 0.65$, $\pi_{ref} = 0.75$, $\Delta_{NI} = 0.30$, $\sigma^2_{center} = 0.20$, $\sigma^2_{center*trt} = 0.005$, $\sigma^2_{block} = 0.15$, $\sigma^2_{pen} = 0.10$

Table 9: *Power for Setting D: Multi-Center, Pen as EU, Binary Outcome, Non-Inferiority*

It should also be noted that for small sample sizes, such as 16 and 40 pens, CI for $\pi_{trt} - \pi_{ref}$ (Independence) and CI for OR were very conservative, generating very low powers compared to CI for $\pi_{trt} - \pi_{ref}$ (Delta Method). But for larger sample sizes of 80 pens, power generated from CI for OR increased more rapidly than that of CI for $\pi_{trt} - \pi_{ref}$ (Independence). It is seen here that even for larger sample sizes, for complex designs, the independence assumption, again, generated wider CIs for difference of proportions which were not independent at all. Also, for small sample sizes, CI for odds was conservative, as expected because of the wider intervals generated during the transformation of limits of the log odds (β). But as the samples became large, the CI performed better, which was again an expected result for a large-sample method such as this [1].

Finally, it should be noted that CI for $\pi_{trt} - \pi_{ref}$ (Delta Method) and CI of odds ratio methods generated similar power results across different blocking designs, while CI for $\pi_{trt} - \pi_{ref}$ (Independence) seemed to be affected by the design. For instance, in settings where there were 4 Center and 80 pens per treatment, power for tests using CI for $\pi_{trt} - \pi_{ref}$ (Independence) had noticeably varying power calculations for CRD (24.5), GRBD(19.1) and RCBD (30.3), while for the other two methods, power seemed to be stable across the different blocking designs. It is, thus, exhibited again that the CI with the independence assumption was unable to take into account the complexity of the design and the consequence of having more random error terms in the model.

4 Discussion and Conclusion

In this report, data simulations, though rigorous and time-consuming, proved to be useful in calculating power of clinical trials with complex designs, and was able to monitor convergence to identify settings that would pose potential convergence issues in the actual data analysis given that such designs would be implemented. The modelling technology of Proc Mixed and Proc Glimmix was used to fit corresponding Linear Mixed Models and Generalized Linear Mixed Models which were chosen as the most appropriate modelling framework for the experimental designs considered.

Compared to the approximation of the non-central F-distribution approach in calculating power described in Verbeke (2000) [24], where an exemplary data set is generated once, the simulation method could additionally monitor convergence rate of mixed models fitted addressing one of the objectives of this project. Furthermore, comparison of confidence limits with the pre-specified margin, which is the standard way of doing non-inferiority testing [23], was not staightforward in the power approximation method especially for binary outcomes, but was easily conducted during simulations once the appropriate confidence intervals were constructed.

Most non-inferiority testing for clinical trials with binary outcomce were done using difference of proportions or odds ratios [23]. To conduct non-inferiority testing then in the simulations, the standard errors of the difference of proportions, $\hat{\pi}_{trt}|\mathbf{b}_i - \hat{\pi}_{ref}|\mathbf{b}_i$, or that of the odds ratio were needed, but the former was not provided by SAS as it would only give out estimates and confidence intervals for parameters in the logit scale and the corresponding mean equivalent using the *ilink* option [18]. For the analyses done with binomial outcomes, these mean estimates then corresponded to the predicted conditional probabilities of success, $\hat{\pi}_{trt}|\mathbf{b}_i = \mathbf{0}$, $\hat{\pi}_{ref}|\mathbf{b}_i = \mathbf{0}$ in the treatment and reference groups, and standard error estimates were only available for them. Thus, two methods were proposed to construct the confidence interval for the difference of proportions to conduct the non-inferiority tests. The first one was by calculating the standard error of the difference using the individual standard error estimates for each proportion, assuming indepence. The second one was by approximating the standard error of the difference through the delta method, assuming that this estimate was asymptotically normal, and that it could be expressed as a function of parameters α and β, which were the estimates in the logit scale with standard error estimates available after fitting a GLMM in Proc Glimmix. The last method for non-inferiority testing was by using the confidence interval constructed for the odds ratio, which was readily available from the mentioned model fitting procedure.

For the simple to moderately complex designs such as Settings A B and C, the first two methods for non-inferiority testing generated very close results, while the tests using CI for OR proved to be very conservative, generating the lowest powers for the settings simulated. This was because the CIs derived were wider due to the enhanced skewness resulting from the transformation of the logit confidence limits [11]. It was also shown in literature that given the same precision level with other parameters such as means and proportions, calculation of sample size based on odds ratio would require a larger sample size [16].

For the most complex design, which was Setting D, wherein simulated trials were multi-centered and pen was the experimental unit, CI for difference of proportions (Independence) generated power estimates which were comparably lower than the other two methods, showing that the assumption of independence was problematic especially for complex designs which employed mixed models with three or more random effects. For all settings considered, CI for difference of proportions (Delta Method) generated the highest power estimates, which was expected since it did not have the constraint of the independence assumption. Furthermore, it was observed that CI for difference of proportions (Delta Method) and CI for OR had stable estimates across the blocking designs, CRD, RCBD and GRBD, implying the ability of the tests to account for the variation between blocks. It should be noted that when the association is not strong, the odds ratio CI would perform well even for small sample sizes.[2] For most of non-inferiority trials then, where usually, the new treatment is not expected to differ with the reference treatment, thus there is no strong association, this CI could be used as a conservative check for power calculated for complex designs.

For the simulations in general, it was observed that when intrablock correlation was larger, power was higher. This was due to the fact that subjects within a block were more correlated, thus, more homogeneous, and the estimation of treatment effect was more efficient[20]. For trials with pen as the experimental unit, however, increasing the intracluster correlation reduced the power of the design. In fact, for some setups simulated where the number of pens were fixed, there was a certain ceiling for power even when the number of animals per pen were greatly increased. This was in agreement with the effective sample size concept and information limit for clustered experiments [8]. For settings with multiple centers, larger between center variability taken away from the residual variance allowed for more efficient estimates [9] , and consequently, generated higher power for the design. However, inclusion of the random center and random center by treatment interaction effects implied that there should be an enough number of centers to estimate these effects. It was exhibited in the simulations that increasing the number of centers greatly elevated the power of the designs. It should be noted that in literature, five centers were even recommended as a minimum number of center to generate stable variance estimates.[3] Moreover, testing of treatment effect for this setting was dependent on the center by treatment interaction variation, thus, this directly affected power estimates. It should be noted that in Table 7, for Scenario 3 where $\sigma^2_{center*trt}$ was increased, and σ^2 was decreased, power was still reduced mainly because the former was still used in testing for the significance of treatment effect. This was not the case for σ^2_{center}, which when accounted for, was totally taken out from the MSE used for the testing.

Results above were expected from the mixed models framework, thus, exhibiting that the simulation method for power calculation was able to mimic experimental designs of interest through its data generation and the significance tests employed. However, knowledge on the random effects were necessary to use this tool. Data from pilot studies should be available, otherwise, appropriately sound estimates should be provided by the clinicians[6]. Furthermore, for simulated designs with small sample sizes, it is possible that the variance components would not be estimated well, hence, it is recommended to check their empirical value from the simulations, and to compare power calculations using the approximation of the non-central F-distribution method[18].

It is also worth noting that the Generalized Linear Mixed Models fitted for the binary outcomes generated estimates which had a conditional interpretation given the random effects. For clinical trials, though, it is the marginal treatment effect of the drug which is of interest. It was, however, pointed out by Agresti (2006) that effect which is significant in the conditional model would usually be significant in the marginal model and the model choice would not influence inferential conclusions on the said parameter. In this report, all testing were limited to the fixed effects parameters in the random effects model.

For most of the simulations, there was no considerable convergence issue encountered. Models used to analyze continuous outcomes generally had 100% convergence rates. The figures were slightly lower for models fitted to binary data but were generally above 90%. It was crucial to monitor this during simulations to verify if the power estimates generated were indeed reliable and were based on a large number of converged models. For the case of some blocked designs in Table 9, where low convergence rates of 85.2%, 83.7% and 89.8% were observed, power estimates generated should be critically considered.

In conclusion, SAS macros used in the simulations were able to calculate power for complex designed veterinary clinical trials considered for which no analytical formulas were readily available. Additionally, it monitored convergence rates, which would be a good measure of the reliability of power estimates generated. This tool is then recommended to be used in future experimental design planning to ensure that sufficient animals would be included in the trial to warrant a power level that would enable the trial to detect the desired magnitude of treatment difference if it indeed exists. In addition, three methods of testing for non-inferiority trials with binary outcomes were proposed. The method using the CI for difference of proportions (Independence) exhibited a certain inability to handle the most complex designs. The method using the CI for odds ratio proved to be conservative in all setups considered. In contrast, the method using the CI for difference of proportions (Delta Method) did not suffer from these limitations. Further investigation on the behavior of these CIs when applied to other experimental settings is recommended.

References

[1] Agresti, A. (2006). *Categorical Data Analysis*. Wiley Series in Probability and Statistics. New Jersey: John Wiley & Sons, Inc.

[2] Agresti, A. (1999). On Logit Confidence Intervals for Odds Ratio with Small Samples. *Biometrics 55,* 597-602.

[3] Brown, H., Prescott, R. (2006). *Applied Mixed Models in Medicine: Second Edition*. Statistics in Practice. West Sussex: John Wiley & Sons, Ltd.

[4] Castelloe, J. (2000). Sample Size Computations and Power Analysis with the SAS System. SAS Institute Inc., Cary, NC. http://www.ats.ucla.edu/stat/sas/library/powersamplesize.pdf. Accessed on 28 August, 2014.

[5] Committee for Medicinal Products for Veterinary use (CVMP) (2012). Guideline on Statistical Principles for Clinical Trials for Veterinary Medicinal Products (Pharmaceuticals). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/01/WC500120834.pdf. Accessed on 24 August, 2014.

[6] Dang, Q., Mazumdar, S., Houck, P. (2008). Sample Size and Power Calculations Based on Generalized Linear Mixed Models with Correlated Binary Outcomes. *Comput Methods Programs Biomed 91(2)*, 122-127.

[7] Duchateau L., Janssen P. and Rowlands G.J. (1998). Linear Mixed Models. An Introduction with Applications in Veterinary Research. ILRI (International Livestock Research Institute), Nairobi, Kenya. http://www.ilri.org/biometrics/Publication/Full%20Text/Linear_Mixed_Models/Toc.htm#P-1_0. Accessed on 26 August, 2014.

[8] Faes, C., Molenberghs, G., Aerts, M., Verbeke, G., Kenward, M. (2009). The Effective Sample Size and an Alternative Small Sample Degrees of Freedom Method. *The American Statistician 63(4)*, 389-399.

[9] Feaster, D., Mikulich-Gilbertson, S., Brincks, A. (2011). Modeling Site Effects in the Design and Analysis of Multisite Trials. *Am J Drug Alcohol Abuse 37(5)*, 383391. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3281513/pdf/nihms-342987.pdf. Accessed on 28 August, 2014.

[10] Galecki, A., Burzykowski, T., Chen, S., Faulkner, J., Ashton-Miller, J. (2009). Statistical Power Calculations for Clustered Continuous Data. *International Journal of Knowledge Engineering and Soft Data Paradigms 1(1)*, 40-48.

[11] Gart, J., Thomas, D. (1982). The Performance of Three Approximate Confidence Limit Methods for the Odds Ratio. *American Journal of Epidemiology (115) 3,* 453-470. http://www.ncbi.nlm.nih.gov/pubmed/7064980. Accessed on 3 September, 2014.

[12] George, S. (2010). Design of Phase III Clinical Trials. *Oncology Clinical Trials*. http://impact.unc.edu/impact7/sites/default/files/OncologyClinicalTrials_Design_of_phase_III_clinical_trials.pdf. Accessed on 31 August, 2014.

[13] Hilton, J. (2010). Non-inferiority Trial Designs for Odds ratios and Risk differences. *Statistics in Medicine 29*, 982-993.

[14] International Cooperation on Harmonization of Technical Requirements for Registration of Veterinary Medicinal Products (VICH) (2000). Good Clinical Practice. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500004343.pdf. Accessed on 24 August, 2014.

[15] Lachin, J. (1981). Introduction to Sample Size Determination and Power Analysis for Clinical Trials. *Controlled Clinical Trials 2*, 93-113.

[16] Lemeshow, S., Hosmer, D., Klar, J., Lwanga, S. (1990). *Adequacy of Sample Size in Health Studies.* World Health Organization. West Sussex: John Wiley & Sons, Ltd.

[17] Lesaffre, E. (2008). Superiority, Equivalence, and Non-inferiority Trials. *Bulletin of the NYU Hospital for Joint Diseases 66(2)*,150-4.

[18] Littell,R., Milliken,G., Stroup, W., Wolfinger, R. and Schabenberger, O. (2006). *SAS for Mixed Models, Second Edition*. Cary, NC: SAS Institute Inc.

[19] Molenberghs, G. &, Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer Series in Statistics. New-York: Springer.

[20] Montgomery, D. (2001). *Design and Analysis of Experiments*. New York: John Wiley & Sons, Inc.

[21] Perrett, J. (2012). A Case Study on Teaching the Topic Experimental Unit and How it is Presented in Advanced Placement Statistics Textbooks. *Journal of Statistics Education 20(2)*. http://www.amstat.org/publications/jse/v20n2/perrett.pdf. 5 September, 2014.

[22] Suissa, S. (1991). Binary Methods for Continuous Outcomes: A Parametric Alternative. *Journal of Clinical Epidemiology, 44(3)* 241-8.

[23] Tunes da Silva, G., Logan , B., Klein, J. (2008). Methods for Equivalence and Noninferiority Testing. *Biology of Blood and Marrow Transplantation : Journal of the American Society for Blood and Marrow Transplantation 15(1 Suppl),* 120-127.

[24] Verbeke, G. &, Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. New-York: Springer.

A Appendix

A.1 Derivation of the Variance of Difference of Proportions (Delta Method)

Variance of π_{ref}

Let $\pi_{ref} = \frac{exp\ \alpha}{1+exp\ \alpha}$, by delta method,

$V(\pi_{ref}) = \left[\frac{\partial \pi_{ref}}{\partial \alpha}\right]^2 V(\alpha)$

$$\begin{aligned}\frac{\partial \pi_{ref}}{\partial \alpha} =& \frac{\partial}{\partial \alpha}\left[\frac{exp\ \alpha}{1+exp\ \alpha}\right]\\ =& \frac{exp\ \alpha}{(1+exp\ \alpha)^2} = \frac{exp\ \alpha}{(1+exp\ \alpha)} * \frac{1}{(1+exp\ \alpha)}\end{aligned}$$

$$\frac{\partial \pi_{ref}}{\partial \alpha} = \pi_{ref}(1-\pi_{ref}). \tag{A1}$$

Thus,

$$V(\pi_{ref}) = \pi_{ref}^2(1-\pi_{ref})^2 V(\alpha). \tag{A2}$$

Variance of π_{trt}

Let $\pi_{trt} = \frac{exp\ (\alpha+\beta)}{1+exp\ (\alpha+\beta)}$, by delta method,

$V(\pi_{trt}) = \left[\frac{\partial \pi_{trt}}{\partial \alpha}\right]^2 V(\alpha) + \left[\frac{\partial \pi_{trt}}{\partial \beta}\right]^2 V(\beta) + 2\frac{\partial \pi_{trt}}{\partial \alpha}\frac{\partial \pi_{trt}}{\partial \beta}Cov(\alpha,\beta)$

$$\begin{aligned}\frac{\partial \pi_{trt}}{\partial \alpha} =& \frac{\partial}{\partial (\alpha)}\left[\frac{exp\ (\alpha+\beta)}{1+exp\ (\alpha+\beta)}\right]\\ =& \frac{exp\ (\alpha+\beta)}{[1+exp\ (\alpha+\beta)]^2}\\ =& \frac{exp\ (\alpha+\beta)}{1+exp\ (\alpha+\beta)} * \frac{1}{1+exp\ (\alpha+\beta)}\end{aligned}$$

$$\frac{\partial \pi_{trt}}{\partial \alpha} = \pi_{trt}(1-\pi_{trt}) \tag{A3}$$

Similarly,

$$\frac{\partial \pi_{trt}}{\partial \beta} = \pi_{trt}(1-\pi_{trt}) \tag{A4}$$

$$\begin{aligned}V(\pi_{trt}) =& \pi_{trt}^2(1-\pi_{trt})^2 V(\alpha) + \pi_{trt}^2(1-\pi_{trt})^2 V(\beta) + 2\pi_{trt}^2(1-\pi_{trt})^2 Cov(\alpha+\beta)\\ =& \pi_{trt}^2(1-\pi_{trt})^2[V(\alpha)+V(\beta)+2Cov(\alpha,\beta)]\end{aligned}$$

Thus,

$$V(\pi_{trt}) = \pi_{trt}^2(1-\pi_{trt})^2 V(\alpha+\beta). \tag{A5}$$

Variance of $\pi_{trt} - \pi_{ref}$

Let $\pi_{trt} - \pi_{ref} = \frac{exp\ (\alpha+\beta)}{1+exp\ (\alpha+\beta)} - \frac{exp\ (\alpha)}{1+exp\ (\alpha)}$, by delta method,

$$V(\pi_{trt}-\pi_{ref}) = \left[\frac{\partial(\pi_{trt}-\pi_{ref})}{\partial\alpha}\right]^2 V(\alpha) + \left[\frac{\partial(\pi_{trt}-\pi_{ref})}{\partial\beta}\right]^2 V(\beta) + 2\frac{\partial(\pi_{trt}-\pi_{ref})}{\partial\alpha}\frac{\partial(\pi_{trt}-\pi_{ref})}{\partial\beta}Cov(\alpha,\beta)$$

$$\begin{aligned}
\frac{\partial(\pi_{trt}-\pi_{ref})}{\partial\alpha} &= \frac{\partial\pi_{trt}}{\partial\alpha} - \frac{\partial\pi_{ref}}{\partial\alpha} \\
&= \pi_{trt}(1-\pi_{trt}) - \pi_{ref}(1-\pi_{ref}) \quad \text{from A1 and A3} \\
\frac{\partial(\pi_{trt}-\pi_{ref})}{\partial\beta} &= \frac{\partial\pi_{trt}}{\partial\beta} - \frac{\partial\pi_{ref}}{\partial\beta} \\
&= \pi_{trt}(1-\pi_{trt}) \qquad\qquad\qquad\quad \text{from A4}
\end{aligned}$$

$$\begin{aligned}
V(\pi_{trt}-\pi_{ref}) =& [\pi_{trt}(1-\pi_{trt}) - \pi_{ref}(1-\pi_{ref})]^2 V(\alpha) + [\pi_{trt}(1-\pi_{trt})]^2 V(\beta) \\
& + 2[\pi_{trt}(1-\pi_{trt}) - \pi_{ref}(1-\pi_{ref})](\pi_{ref}(1-\pi_{ref})Cov(\alpha,\beta) \\
\\
=& \pi_{trt}^2(1-\pi_{trt})^2 V(\alpha) + \pi_{trt}^2(1-\pi_{trt})^2 V(\beta) + 2\pi_{trt}^2(1-\pi_{trt})^2 Cov(\alpha,\beta) \\
& + \pi_{ref}^2(1-\pi_{ref})^2 V(\alpha) \\
& - 2\pi_{trt}(1-\pi_{trt})\pi_{ref}(1-\pi_{ref})V(\alpha) - 2\pi_{trt}(1-\pi_{trt})\pi_{ref}(1-\pi_{ref})Cov(\alpha,\beta) \\
\\
=& \pi_{trt}^2(1-\pi_{trt})^2 V(\alpha+\beta) + \pi_{ref}^2(1-\pi_{ref})^2 V(\alpha) \\
& - 2\pi_{trt}(1-\pi_{trt})\pi_{ref}(1-\pi_{ref})V(\alpha) - 2\pi_{trt}(1-\pi_{trt})\pi_{ref}(1-\pi_{ref})Cov(\alpha,\beta)
\end{aligned}$$

Thus,

$$V(\pi_{trt}-\pi_{ref}) = V(\pi_{trt}) + V(\pi_{ref}) - 2\pi_{trt}(1-\pi_{trt})\pi_{ref}(1-\pi_{ref})[V(\alpha) + Cov(\alpha,\beta)] \qquad \text{(A6)}$$

from A2 and A5.

A.2 Auxiliary Results

A.2.1 Superiority Testing - Continuous Outcome

Parameter	RCBD/ GRBD		CRD	
σ^2_{center}	0.04	9%	0.04	13%
$\sigma^2_{center*trt}$	0.01	2%	0.01	3%
σ^2_{pen}	0.15	33%	0.15	50%
σ^2_{block}	0.15	33%		0%
σ^2_{res}	0.10	22%	0.10	33%
σ^2_{total}	0.45	100%	0.30	100%
σ_{total}	0.67		0.55	
Δ_{abs}	0.54		0.54	
$\Delta_{std} = \frac{\Delta}{\sigma_{total}}$	0.80		0.99	

Table A1: *Pre-specified Parameters in Setting D: Multi-Center, Pen as EU, Continuous Outcome*

Design	Centers	Blocks/ Center	Pens/ Block/ Center/ Trt	Animals/ Pen/ Block/ Center/ Trt	Total Animals/ Trt	Total Pens/ Trt	Conver	Power
CRD	4	-	4	2	32	16	100.0	46.7
	4	-	10	2	80	40	100.0	83.2
	4	-	20	2	160	80	100.0	94.4
	8	-	10	2	160	80	100.0	100.0
GRBD	4	2	2	2	32	16	100.0	44.9
	4	2	5	2	80	40	100.0	88.0
	4	2	10	2	160	80	100.0	97.5
	8	2	5	2	160	80	100.0	100.0
RCBD	4	4	1	2	32	16	100.0	60.1
	4	10	1	2	80	40	100.0	93.0
	4	20	1	2	160	80	100.0	99.7
	8	10	1	2	160	80	100.0	100.0

Table A2: *Power for Setting D: Multi-Center, Pen as EU, Continuous Outcome, Superiority*

A.2.2 Superiority Testing - Binary Outcome

Design	Blocks	Animal/ Block/ Treatment	Animal/ Treatment	Convergence (%)	Power
CRD	0	12	12	100.0	18.2
		20	20	100.0	44.1
		40	40	100.0	82.4
		80	80	100.0	98.0
GRBD	3	4	12	96.4	20.0
	5	4	20	99.7	44.5
	10	4	40	99.5	79.1
	20	4	80	99.2	96.9
RCBD	12	1	12	97.6	19.1
	20	1	20	99.9	40.6
	40	1	40	100.0	77.8
	80	1	80	100.0	97.3

$\pi_{trt} = 0.75$, $\pi_{ref} = 0.45$, $\sigma^2_{block} = 0.15$

Table A3: *Power for Setting A: Single Center, Animal as EU, Binary Outcome, Superiority*

Design	Blocks	Pen/ Block/ Treatment	Animal/ Pen/ Block/ Treatment	Pen/ Treatment	Animal/ Treatment	Convergence (%)	Power
CRD	-	2	8	2	16	98.9	0.0
	-	2	16	2	32	99.4	0.5
	-	2	40	2	80	99.3	19.1
	-	2	80	2	160	98.7	38.1
	-	8	2	8	16	99.1	22.6
	-	16	2	16	32	100.0	59.4
	-	40	2	40	80	100.0	96.3
	-	80	2	80	160	98.9	100.0
GRBD	2	4	2	8	16	98.8	21.0
	2	8	2	16	32	99.2	61.0
	2	20	2	40	80	98.7	96.7
	2	40	2	80	160	98.0	100.0
RCBD	8	1	2	8	16	99.5	20.0
	16	1	2	16	32	99.9	57.4
	40	1	2	40	80	99.0	96.4
	80	1	2	80	160	98.7	100.0

$\pi_{trt} = 0.75$, $\pi_{ref} = 0.45$, $\sigma^2_{block} = 0.15$, $\sigma^2_{pen} = 0.10$

Table A4: *Power for Setting B: Single Center, Pen as EU, Binary Outcome, Superiority*

Design	Centers	Block/ Center/ Treatment	Animals/ Block/ Treatment	Animals/ Center/ Treatment	Animals/ Treatment	Convergence (%)	Power
CRD(*)	3	0	4	4	12	96.6	0.0
	5	0	4	4	20	99.2	10.6
	10	0	4	4	40	99.7	62.6
	20	0	4	4	80	99.0	95.6
GRBD	3	2	2	4	12	96.2	0.0
	5	2	2	4	20	99.8	10.6
	10	2	2	4	40	99.3	63.3
	20	2	2	4	80	98.8	94.1
RCBD(*)	3	4	1	4	12	97.4	0.0
	5	4	1	4	20	99.5	8.6
	10	4	1	4	40	100.0	60.0
	20	4	1	4	80	99.2	94.3

$\pi_{trt} = 0.75$, $\pi_{ref} = 0.45$, $\sigma^2_{block} = 0.15$, $\sigma^2_{center} = 0.20$, $\sigma^2_{center*trt} = 0.005$

Table A5: *Power for Setting C: Multi-Center, Animal as EU, Binary Outcome, Superiority*

Design	Centers	Blocks/ Center	Pens/ Block/ Center/ Trt	Animals/ Pen/ Block/ Center/ Trt	Total Animals/ Trt	Total Pens/ Trt	Conver	Power
CRD	4	-	4	2	32	16	99.8	9.5
	4	-	10	2	80	40	94.7	54.2
	4	-	20	2	160	80	92.2	89.4
	8	-	10	2	160	80	93.3	99.2
GRBD	4	2	2	2	32	16	99.1	9.2
	4	2	5	2	80	40	95.3	55.5
	4	2	10	2	160	80	87.5	89.7
	8	2	5	2	160	80	94.0	99.1
RCBD	4	4	1	2	32	16	99.7	10.2
	4	10	1	2	80	40	93.4	54.4
	4	20	1	2	160	80	84.7	89.4
	8	10	1	2	160	80	89.1	99.6

$\pi_{trt} = 0.75$, $\pi_{ref} = 0.45$, $\sigma^2_{block} = 0.15$, $\sigma^2_{pen} = 0.10$, $\sigma^2_{center} = 0.20$, $\sigma^2_{center*trt} = 0.005$

Table A6: *Power for Setting D: Multi-Center, Pen as EU, Binary Outcome, Superiority*

A.2.3 Non-Inferiority Testing - Binary Outcome

							Power		
Design	Blocks	Pens/ Block/ Trt	Animals/ Pen/ Block/ Trt	Total Pens/ Trt	Total Animals/ Trt	Convergence (%)	$\pi_{trt} - \pi_{ref}$ (Indep)	$\pi_{trt} - \pi_{ref}$ (Delta)	Odds Ratio
CRD	-	2	8	2	16	98.4	0.2	0.2	0.0
	-	2	16	2	32	99.6	0.8	0.8	0.4
	-	2	40	2	80	99.4	7.1	7.1	3.9
	-	2	80	2	160	98.3	21.0	21.0	12.3
	-	8	2	8	16	98.4	19.3	19.3	11.8
	-	16	2	16	32	100.0	36.4	36.4	25.2
	-	40	2	40	80	100.0	77.5	77.5	61.6
	-	80	2	80	160	97.9	97.1	97.1	89.1
GRBD	2	4	2	8	16	98.7	12.2	16.9	8.5
	2	8	2	16	32	99.2	25.7	34.2	24.1
	2	20	2	40	80	97.9	53.4	74.3	58.4
	2	40	2	80	160	97.4	71.0	96.3	90.5
RCBD	8	1	2	8	16	99.2	12.3	14.7	8.6
	16	1	2	16	32	100.0	30.8	33.6	23.6
	40	1	2	40	80	99.3	74.8	76.6	61.8
	80	1	2	80	160	99.2	96.6	97.4	90.8

$\pi_{trt} = 0.65$, $\pi_{ref} = 0.75$, $\Delta_{NI} = 0.30$,$\sigma^2_{block} = 0.15$, $\sigma^2_{pen} = 0.10$

Table A7: *Power for Setting B: Single Center, Pen as EU, Binary Outcome, Non-Inferiority*

							Power		
Design	Centers	Blocks/ Center	Animals/ Block/ Center/ Trt	Total Animals/ Center/ Trt	Total Animals/ Trt	Convergence(%)	$\pi_{trt} - \pi_{ref}$ (Indep)	$\pi_{trt} - \pi_{ref}$ (Delta)	Odds Ratio
CRD	3	-	4	4	12	95.6	0.0	0.0	0.0
	5	-	4	4	20	99.4	6.0	7.9	4.0
	10	-	4	4	40	99.9	30.9	36.8	24.5
	20	-	4	4	80	98.8	68.3	72.8	57.6
GRBD	3	2	2	4	12	96.1	0.0	0.0	0.0
	5	2	2	4	20	99.2	5.1	7.3	3.8
	10	2	2	4	40	99.9	26.4	33.6	22.6
	20	2	2	4	80	99.1	64.3	71.8	57.2
RCBD	3	4	1	4	12	95.9	0.0	0.0	0.0
	5	4	1	4	20	99.7	6.8	9.3	4.8
	10	4	1	4	40	99.7	30.5	36.4	25.8
	20	4	1	4	80	99.3	69.5	74.9	60.0

$\pi_{trt} = 0.65$, $\pi_{ref} = 0.75$, $\Delta_{NI} = 0.30$, $\sigma^2_{block} = 0.15$, $\sigma^2_{center} = 0.20$, $\sigma^2_{center*trt} = 0.005$

Table A8: *Power for Setting C: Multi-Center, Animal as EU, Binary Outcome, Non-Inferiority*

A.3 Sample Macro Codes

A.3.1 Single Center, Animal as EU, Continuous Outcome, Superiority

```
**** Single site, Continuous data, Animal experimental unit and Superiority;

%let alpha=.05;         **** sig level (two-sided);
%let seed1=260288; **** random block;
%let seed2=160288;     **** random response;
%let nsim=1000;         **** number of simulated data sets;
%let mu=14.5;       **** mean in ref (T02) group;
%let nu=14.75;      **** mean in trt (T01) group;
%let vb=0.1500; **** block variability;  /*for CRD, this is fixed to 0*/
%let vr=0.2500;     **** residual variability;

%macro mpwr(design=, nblock=, nanim= );

/*Generating data and fitting a model for CRD*/
***nblock= number of blocks per treatment;
***nanim= number of animals per block/treatment;

********************************************************************************;
***Generating data and fitting a model for CRD;
********************************************************************************;
***CRD: Design has no blocks. Animals are randomized to Treatments;
***Even if numbers are entered for nblock, it is not used in generating the data;
***nanim= number of animals per treatment;

%if &design="CRD" or &design="crd" %then
%do;
data cult60;
   mu=&mu;
   nu=&nu;
   vr=&vr;
do sim=1 to &nsim;
do trt = 1 to 2;
     if (trt=1) then mean=nu;
        else  mean=mu;
      do animal=1 to &nanim;
         y=mean+sqrt(vr)*rannor(&seed2);
output;
     end;
end;
```

```
end;

keep nu mu sim trt animal y;
run;

proc sort data=cult60;
   by nu mu sim trt;
run;

ods listing exclude all;
proc mixed data=cult60;
by nu mu sim ;
class trt;
  model y = trt/solution;
  contrast ’1 Vs 2’ trt 1 -1;
  ods output contrasts=c; ***Convergence Status is not reported.
   ***No iterations done for this setting;
run;
ods listing exclude none;
%end;


*******************************************************************;
***Generating data and fitting a model for RCBD;
*******************************************************************;
***RCBD: Design has blocks, and each block contains only 1 animal. ;
***Even if numbers are entered for nanim, macro sets nanim=1;

%if &design="RCBD" or &design="rcbd" %then
%do;
data cult60;
   mu=&mu;
   nu=&nu;
   vb=&vb;
   vr=&vr;
   nanim=1; /*fixed by design - RCBD*/

do sim=1 to &nsim;
do block=1 to &nblock;
zblock=rannor(&seed1)*sqrt(vb);

do trt = 1 to 2;
         if (trt=1) then mean=nu;
          else  mean=mu;
       do animal=1 to nanim;
        y=mean+zblock+sqrt(vr)*rannor(&seed2);
           output;
       end;
     end;
end;
end;

keep nu mu sim trt block animal y;
run;

proc sort data=cult60;
   by nu mu  sim trt block;
run;

ods listing exclude all;
proc mixed data=cult60;
by nu mu  sim ;
class trt block;
model y = trt/solution;
   random block;
contrast ’1 Vs 2’ trt 1 -1 ;
ods output   ConvergenceStatus=status contrasts=c;
run;
ods listing exclude none;
%end;


*******************************************************************;
```

```
***Generating data and fitting a model for GRBD;
****************************************************************;
***GRBD: Design has blocks, and each block can contain multiple animals;

%if &design="GRBD" or &design="grbd" %then
%do;
data cult60;
   mu=&mu;
   nu=&nu;
   vb=&vb;
   vr=&vr;
do sim=1 to &nsim;
do block=1 to &nblock;
zblock=rannor(&seed1)*sqrt(vb);

   do trt = 1 to 2;
         if (trt=1) then mean=nu;
         else  mean=mu;
      do animal=1 to &nanim;
    y=mean+zblock+sqrt(vr)*rannor(&seed2);
          output;
      end;
    end;
end;
    end;
  keep nu mu sim trt block animal y;
run;

proc sort data=cult60;
   by nu mu  sim trt block;
run;

ods listing exclude all;
proc mixed data=cult60;
  by nu mu  sim ;
  class trt block;
   model y = trt/solution;
   random block;
contrast '1 Vs 2' trt 1 -1;
ods output   ConvergenceStatus=status contrasts=c;
run;
ods listing exclude none;
%end;

/*Calculating convergence rate*/
data converge;
set status;
by nu mu sim;
length conv $3;
   if Status=0 then conv='yes';
   else conv='no';
   nsim=&nsim;
run;
/*Calculating  power - through percentage of significant results*/
data signi;
set c;
by nu mu sim;
length sig $3;
alpha=&alpha;
  if probf<=alpha then sig='YES';
  else sig='NO';
  nsim=&nsim;
run;


/*Extracting necessary results*/

/*Results for convergence*/
proc freq data=converge noprint;
   by nu mu;
   table conv/list out=outconv;
run;
```

```
***For CRD, Convergence Status data set is not produced
***since this model has no random effect, and iteration is not necessary.;
***Convergence rate is set to missing;
%if &design="CRD" or &design="crd" %then
%do;
data outconve;
nu=&nu;
mu=&mu;
percent1=.;
run;
%end;

%if &design="RCBD" or &design="rcbd" or &design="GRBD" or &design="grbd" %then
%do;
data outconve;
set outconv;
where conv=’yes’;
percent1=percent;
run;
%end;

/*Results for power*/
proc freq data=signi noprint;
   by nu mu;
   table sig/list out=outsig;
run;

data outsign;
set outsig;
where sig=’YES’;
percent2=percent;
run;

/*Merging all results*/
data outall;
merge outSIGN outconve;
by nu mu;
length design $8;

if &design="CRD" or &design="crd" then
do;
nblock=0;
nanim=&nanim;
design=’CRD(*)’;
tanim=&nanim; *number of animals per treatment;
end;

if &design=’RCBD’ or &design=’rcbd’ then
do;
nblock=&nblock;
nanim=1;
design=’RCBD(*)’;
tanim=&nblock; *number of animals per treatment;
end;

if &design=’GRBD’ or &design=’grbd’ then
do;
nblock=&nblock;
nanim=&nanim;
design=’GRBD’;
tanim=&nblock*&nanim; *number of animals per treatment;
end;
run;

/*Final Output*/

***For CRD, percent1 was automatically set to missing as convergence status is not applicable;
***However, calculated power is still valid, thus it is included in the output data set;

%if &design="CRD" or &design="crd"
%then
%do;
data outalll;
```

```
set outall;
run;
proc append base=results data=outalll(keep=design nu mu nblock nanim tanim percent1 percent2);
run;
%end;

%else %if &design="RCBD" or &design="GRBD" or &design="rcbd" or &design="grbd"
%then
%do;

***For RCBD and GRBD, when percent1 is missing, there is no convergence thus
calculated power is not valid, and is set to missing;
data outalll;
set outall;
if percent1=. then percent1=.;
if percent1=. then percent2=.;
run;

proc append base=results data=outalll(keep=design nu mu nblock nanim tanim percent1 percent2);
run;
%end;

/*Setting values to null when specified design is not correct*/
%else  %do;
data outalll;

design=’invalid ’;
nu=.;
mu=.;
nblock=.;
nanim=.;
tanim=.;
percent1=.;
percent2=.;

putlog "Error: Invalid Design Type";
run;

proc append base=results data=outalll(keep=design nu mu nblock nanim tanim percent1 percent2);
run;  ***deletes all files except macro and compiled results;
%end;

proc delete data=c converge cult60 outall outalll outconv outconve outsig outsign signi status
(gennum=all);
run;
quit;

%mend mpwr;


%mpwr(design="CRD", nblock=0, nanim=12);  **nblock is 0 for CRD ;
%mpwr(design="crd", nblock=0, nanim=20);  **nanim is the number of animals per treatment;
%mpwr(design="crd", nblock=0, nanim=40);

%mpwr(design="RCBD", nblock=12, nanim=1); **nblock is the number of blocks per treatment;
%mpwr(design="rcbd", nblock=20, nanim=1); **nanim is not needed for RCBD design: default value:1;
%mpwr(design="rcbd", nblock=40, nanim=1);

%mpwr(design="GRBD", nblock=3, nanim=4); **nblock is the number of blocks per treatment;
%mpwr(design="grbd", nblock=5, nanim=4); **nanim is the number of animals per block/treatment;
%mpwr(design="grbd", nblock=10, nanim=4);

title;
footnote;
************************************************************************************************;
title1 height=10pt "Power Estimates - Superiority for T01 (Treatment) VS T02 (Reference)";
title2 height=10pt "Number of Treatments=2, Alpha=&alpha (two-sided), Number of Simulations= &nsim";
title3 height=10pt "Single Site; Animal as EU; Continuous Response";

************************************************************************************************;
options missing=’ ’ nocenter;
proc report data=results nowindows split=’*’ spacing=1 headline missing formchar(2)="-";
   column (’Means’ nu mu) design nblock nanim tanim percent1 percent2;
   define nu      / order format=8.2 ’T01 (Trt)’;
   define mu     / order format=8.2 ’T02 (Ref)’;
   define design    / order format=$8. ’Design’;
   define nblock    / order format=8. ’Number of Blocks/ Trt’;
   define nanim    / order format=8. ’Number of Animals/ Block/ Trt’ ;
   define tanim    / order format=8. ’Total Number of Animals/ Trt’ ;
   define percent1  / display format=8.1 ’Conver (%)’;
   define percent2  / display format=8.1 ’Power (%)’;

footnote1 j=left height=8pt "Variability:";
footnote2 j=left height=8pt "    Block   = &vb";
footnote3 j=left height=8pt "    Residual   = &vr";
footnote4 ;
footnote5 j=left height=8pt "CRD(*)  1. Number of Animals/blk/trt = Number of Animals/trt";
footnote6 j=left height=8pt "        2. Block Variability = 0";
footnote7 j=left height=8pt "        3. Convergence does not apply to CRD Setting.";
footnote8 ;
footnote9 j=left height=8pt "RCBD(*) Number of Animals/blk/trt = 1 by default";
run;
quit;
```

A.3.2 Multi-Center, Pen as EU, Binary Outcome, Non-Inferiority

```
%let alpha=.025;        **** sig level (one-sided);
%let alpha_ci=.05;    **** sig level*2 (will be used by SAS to determine LCL for OR);
%let seed1=2848377; **** random site;
%let seed2=3602883; **** random block;
%let seed3=3216329; **** random site*treatment;
%let seed4=7797318;     **** random pen;
%let seed5=2984382; **** random response;
%let nsim=1000;         **** number of simulated data sets;
%let ref=0.75;    **** cure rate in (T02) group;
%let new=0.65;    **** cure rate in (T01) group;
%let delta=0.3;      **** clinically acceptable difference;
%let vs=0.2000;    **** site  variability(logit scale);
%let vst=0.0050;   **** site by treatment variability(logit scale);
%let vp=0.1000;     **** pen variability(logit scale);
%let vb=0.1500;  **** block variability(logit scale); /*for CRD, this is fixed to 0*/


%macro mpwr(nsite=,nblock=,npen=, nanim=, design= );
***nsite= number of sites;
***nblock= number of blocks per/site/treatment;
***npen= number of pens per block/site/treatment;
***nanim= number of animals per pen/block/site/treatment;

*******************************************************************************;
***Generating data and fitting a model for CRD;
*******************************************************************************;
***CRD: Design has no blocks. Pens are randomized to Treatments;
***Even if numbers are entered for nblock, it is not used in generating the data;
***nanim= number of animals per pen/site/treatment;

%if &design="CRD" or &design="crd" %then
%do;
data cure;
   ref=&ref;
   new=&new;
   delta=&delta;
    vs=&vs;
    vst=&vst;
    vp=&vp;

   mu=log(ref/(1-ref));
   nu=log(new/(1-new));

   do sim=1 to &nsim;
   do site=1 to &nsite;
    zsite=rannor(&seed1)*sqrt(vs);
```

```
     do trt = 1 to 2;
       zst=rannor(&seed3)*sqrt(vst);
             if (trt eq 1) then mean=nu;
                else mean=mu;

do pen=1 to &npen;
zpen=rannor(&seed4)*sqrt(vp);

             do animal = 1 to &nanim;
                phat=exp(mean+zsite+zst+zpen)/(1+exp(mean+zsite+zst+zpen));
y=ranbin(&seed5,1,phat);
               output;
 end; end; end; end; end;
   keep delta ref new site pen trt animal sim y;
run;

proc sort data=cure;
   by  delta ref new sim site trt pen;
run;

ods listing exclude all;
proc glimmix data=cure;
   by delta ref new sim;
   class trt pen site;
   model y = trt / link= logit dist=bin alpha=&alpha_ci oddsratio covb;
   random site site*trt pen(site*trt);
  lsmeans trt/cl ilink;
   ods output   ConvergenceStatus=status lsmeans=lsmeans OddsRatios=OR covB=cov;
run;
ods listing exclude none;
%end;

*******************************************************************;
***Generating data and fitting a model for RCBD;
*******************************************************************;
***RCBD: Design has blocks, and each block contains only 1 pen. ;
***Even if numbers are entered for npen, macro sets npen=1;

%if &design="RCBD" or &design="rcbd" %then
%do;
data cure;
   ref=&ref;
   new=&new;
   delta=&delta;
    vs=&vs;
    vst=&vst;
vb=&vb;
    vp=&vp;
npen=1; /*fixed by design - RCBD*/

   mu=log(ref/(1-ref));
   nu=log(new/(1-new));

   do sim=1 to &nsim;
   do site=1 to &nsite;
    zsite=rannor(&seed1)*sqrt(vs);

    do block= 1 to &nblock;
zblock=rannor(&seed2)*sqrt(vb);

     do trt = 1 to 2;
       zst=rannor(&seed3)*sqrt(vst);
             if (trt eq 1) then mean=nu;
                else mean=mu;

do pen=1 to npen;
zpen=rannor(&seed4)*sqrt(vp);

             do animal = 1 to &nanim;
                phat=exp(mean+zsite+zblock+zst+zpen)/(1+exp(mean+zsite+zblock+zst+zpen));
y=ranbin(&seed5,1,phat);
               output;
 end; end; end; end; end; end;
```

```
   keep delta ref new site block pen trt animal sim y;
run;

proc sort data=cure;
   by  delta ref new sim site block trt pen  ;
run;

ods listing exclude all;
proc glimmix data=cure;
   by delta ref new sim;
   class trt pen block site;
   model y = trt / link= logit dist=bin alpha=&alpha_ci oddsratio covb;
   random site block(site) site*trt trt*block*site;
  lsmeans trt/cl ilink;
   ods output   ConvergenceStatus=status lsmeans=lsmeans OddsRatios=OR covB=cov;
run;
ods listing exclude none;
%end;

******************************************************************;
***Generating data and fitting a model for GRBD;
******************************************************************;
***GRBD: Design has blocks, and each block can contain multiple pens;

%if &design="GRBD" or &design="grbd" %then
%do;
data cure;
   ref=&ref;
   new=&new;
   delta=&delta;
    vs=&vs;
    vst=&vst;
vb=&vb;
    vp=&vp;

   mu=log(ref/(1-ref));
   nu=log(new/(1-new));

   do sim=1 to &nsim;
   do site=1 to &nsite;
    zsite=rannor(&seed1)*sqrt(vs);

    do block= 1 to &nblock;
zblock=rannor(&seed2)*sqrt(vb);

     do trt = 1 to 2;
       zst=rannor(&seed3)*sqrt(vst);
             if (trt eq 1) then mean=nu;
                else mean=mu;

do pen=1 to &npen;
zpen=rannor(&seed4)*sqrt(vp);

             do animal = 1 to &nanim;
                phat=exp(mean+zsite+zblock+zst+zpen)/(1+exp(mean+zsite+zblock+zst+zpen));
y=ranbin(&seed5,1,phat);
               output;
 end; end; end; end; end; end;
   keep delta ref new site block pen trt animal sim y;
run;

proc sort data=cure;
   by  delta ref new sim site block trt pen  ;
run;

ods listing exclude all;
proc glimmix data=cure;
   by delta ref new sim;
   class trt pen block site;
   model y = trt / link= logit dist=bin alpha=&alpha_ci oddsratio covb;
   random site block(site) site*trt pen(block*site*trt);
   lsmeans trt/cl ilink;
   ods output   ConvergenceStatus=status lsmeans=lsmeans OddsRatios=OR covB=cov;
```

```
run;
ods listing exclude none;
%end;

/*Calculating Convergence Rate*/
data converge;
   set status;
   by delta ref new sim;
   length conv $3;
   if Status=0 then conv=’yes’;
   else conv=’no’;
   nsim=&nsim;
run;

/*Calculating power from simulations*/
data lsmeans_1;
   set lsmeans;
   if trt=1;
   mu1=mu;
   se1=stderrmu;
   df1=df;
   keep delta ref new sim mu1 se1 df1;
run;

data lsmeans_2;
   set lsmeans;
   if trt=2;
   mu2=mu;
   se2=stderrmu;
   df2=df;
   keep delta ref new  sim mu2 se2 df2;
run;

/*Non-inferiority CI for difference of prop using the SE calculated assuming independence*/
/*CI Construction using t-stat with df equal to that of test for fixed effect*/
data Signi_ind;
   merge lsmeans_1 lsmeans_2;
   by delta ref new  sim;
   length non $3;

   ng_delta=-1*delta;
   diff = mu1 - mu2;
   mse_hat_d = sqrt(se1**2 + se2**2);       **** Assumes independence;
   t = tinv(1-&alpha,df1); **** using df of estimate for treatment diff in logit scale;
   lower_t = diff - t*mse_hat_d; **** LCL of the difference of prop;
   if lower_t gt ng_delta then non = ’yes’;  **** Non-inferior if LCL > - delta;
   else if lower_t le ng_delta then non = ’no’;
   nsim=&nsim;
run;

/*Non-inferiority CI for difference of prop using the SE calculated through Delta Method*/
/*Extracting variance of parameters in the logit scale*/
data cov_parms;
set cov;
by delta ref new sim;
   var_alpha=col1;
   cov_alpha_beta=col2;
keep delta ref new sim var_alpha cov_alpha_beta ;
where effect="Intercept";
run;

/*CI Construction using t-stat with df equal to that of test for fixed effect*/
data signi_delta;
merge lsmeans_1 lsmeans_2 cov_parms;
   by delta ref new sim;

diff=mu1-mu2;
se_diff=sqrt(se1**2+se2**2-(2*mu1*(1-mu1)*mu2*(1-mu2)*(var_alpha+cov_alpha_beta)));
/*SE of difference of prop- DELTA METHOD APPROXIMATION*/
    ng_delta=-1*delta;
    t = tinv(1-&alpha,df1); **** using df of estimate for treatment diff in logit scale;
    lower_t = diff - t*se_diff; **** LCL of the difference of prop;
    if lower_t gt ng_delta then non = ’yes’;    **** Non-inferior if LCL > - delta;
    else if lower_t le ng_delta then non = 'no';
    nsim=&nsim;
run;

/*Non-inferiority CI for Odds Ratio*/
data Signi_OR;
   set OR;
   by delta ref new sim;
   length non $3;

   or_delta= ((ref-delta) /(1-(ref-delta)))/(ref /(1-ref));
    ***expressing the non-inferiority margin in the OR scale;

if or_delta >= 0 then
    do;
    if lower gt or_delta then non = 'yes';    **** Non-inferior if OR_LCL >  OR_delta (NI margin);
    else if lower le or_delta then non = 'no';
    nsim=&nsim;
    end;

   ***OR value should lie from zero to infinity;
   ***Significance testing is not done using OR when delta used causes
   the OR non-inferiority margin to have negative values;
   else if or_delta < 0 then
   or_delta=.;
   nsim=&nsim;
run;


/*Extracting necessary results*/

/*Results for convergence*/
proc freq data=converge noprint;
   by delta ref new ;
   table conv/list out=outconv;
run;

data outconve;
set outconv;
where conv='yes';
percent1=percent;
run;

/*Results for power - using diff of prop (independence assumption)*/
proc freq data=signi_ind noprint;
   by delta ref new;
   table non /list out=outsig_ind;
run;

data outsign_ind;
set outsig_ind;
where non='yes';
percent2=percent;
run;


/*Results for power - using diff of prop (delta method)*/
proc freq data=signi_delta noprint;
   by delta ref new;
   table non /list out=outsig_delta;
run;

data outsign_delta;
set outsig_delta;
where non='yes';
percent3=percent;
run;

/*Results for power - using Odds Ratio CI*/
proc freq data=signi_or noprint;
   by delta ref new;
   table non /list out=outsig_or;
run;
```

```
data outsign_or;
set outsig_or;
where non='yes';
percent4=percent;
run;

/*Merging all results*/
data outall;
merge outsign_ind outsign_delta outsign_or outconve;
by delta ref new;
length design $8;

if &design="CRD" or &design="crd" then
do;
nsite=&nsite;
nblock=0;
npen=&npen;
nanim=&nanim;
design='CRD(*)';
tanim_site=&nanim*&npen; *number of animals per site/treatment;
tanim=&nsite*&npen*&nanim; *number of animals per treatment;
    tpen_site=&npen; *number of pens per site/treatment;
tpen=&nsite*&npen; *number of pens per treatment;
end;

if &design='RCBD' or &design='rcbd' then
do;
nsite=&nsite;
nblock=&nblock;
npen=1;
nanim=&nanim;
design='RCBD(*)';
tanim_site=&nanim*&nblock; *number of animals per site/treatment;
tanim=&nsite*&nblock*&nanim; *number of animals per treatment;
tpen_site=&nblock; *number of pens per site/treatment;
tpen=&nsite*&nblock; *number of pens per treatment;
end;
if &design='GRBD' or &design='grbd' then
do;
nsite=&nsite;
nblock=&nblock;
npen=&npen;
nanim=&nanim;
design='GRBD';
tanim_site=&nanim*&nblock*&npen; *number of animals per site/treatment;
tanim=&nsite*&nblock*&npen*&nanim; *number of animals per treatment;
tpen_site=&nblock*&npen; *number of pens per site/treatment;
tpen=&nsite*&nblock*&npen; *number of pens per treatment;
end;
run;

/*Final Output*/
%if &design="CRD" or &design="RCBD" or &design="GRBD" or &design="crd"
or &design="rcbd" or &design="grbd"
%then
%do;

/*Setting values to missing when convergence is never achieved*/

data outalll;
set outall;
if percent1=. then percent1=.;
if percent1=. then percent2=.;
if percent1=. then percent3=.;
if percent1=. then percent4=.;
run;

proc append base=results data=outalll(keep=delta ref new design nsite nblock npen
nanim tanim_site tanim tpen_site tpen percent1 percent2 percent3 percent4);
run;
%end;
```

```
/*Setting values to null when specified design is not correct*/
%else  %do;
data outalll;

delta=.; ref=.; new=.; design='invalid '; nsite=.; nblock=.; npen=.; nanim=.;
tanim=.; tanim_site=.; tpen_site=.; tpen=.; percent1=.; percent2=.;
percent3=.; percent4=.; run;

proc append base=results data=outalll(keep=delta ref new design nsite nblock npen
nanim tanim tanim_site tpen_site tpen percent1 percent2 percent3  percent4);
run;
%end;

proc delete data=converge cov cov_parms cure lsmeans lsmeans_1 lsmeans_2
or outall outalll outconv outconve outsign_delta outsign_ind outsign_or
outsig_delta outsig_ind outsig_or signi_delta signi_ind signi_or status (gennum=all);
run; ***deletes all files except macro and compiled results;
quit;

%mend mpwr;

%mpwr(design="crd", nsite=4, nblock=0, npen=2,nanim=4); **nsite is the number of sites;
%mpwr(design="CRD", nsite=4, nblock=0, npen=2,nanim=10);**nblock is 0 for CRD ;
%mpwr(design="crd", nsite=4, nblock=0, npen=2,nanim=20);**npen is the number of pens/site/trt;

%mpwr(design="crd", nsite=4, nblock=0, npen=4,nanim=2);
**nanim is the number of animals/pen/site/trt;

%mpwr(design="rcbd", nsite=4, nblock=4, npen=1,nanim=2);**nblock is the number of blocks/site/trt;

%mpwr(design="RCBD", nsite=4, nblock=10, npen=1,nanim=2);
**npen is not needed for RCBD: default value is 1;

%mpwr(design="rcbd", nsite=4, nblock=20, npen=1,nanim=2);
**nanim is the number of animals/pen/block/site/trt;


%mpwr(design="grbd", nsite=4, nblock=2, npen=2,nanim=2);**nblock is the number of blocks/site/trt;
%mpwr(design="GRBD", nsite=4, nblock=2, npen=5,nanim=2);
**npen is the number of pens/block/site/trt;

%mpwr(design="grbd", nsite=4, nblock=2, npen=10,nanim=2);
**nanim is the number of animals/pen/block/site/trt;

title;
footnote;
*******************************************************************************************;
title1 "Power Estimates  - Non-inferiority for T01 (Treatment) VS T02 (Reference)";
title2 "Number of Treatments = 2, Alpha = &alpha (one-sided), Number of Simulations= &nsim";
title3 "Multiple Site; Pen as EU; Binary Response";
*******************************************************************************************;

options missing=' ' nocenter orientation=landscape;
proc report data=results nowindows split='*' spacing=1 headline missing formchar(2)="-";
   column delta ('Cure Rates' new ref) design nsite nblock npen nanim tanim_site tanim
                                       tpen_site tpen
   percent1 ('Power'('Diff of Prop' percent2 percent3 ) percent4);
   define delta   / order format=5.2 'Delta';
   define new    / order format=5.2 ' T01 (Trt)';
   define ref    / order format=5.2 'T02 (Ref)';
   define design    / order format=$7. 'Design';
   define nsite    / order format=6. 'Number of Sites';
   define nblock    / order format=7. 'Number of Blocks/ Site/ Trt';
   define npen    / order format=6. 'Number of Pens/ Blocks/ Site/ Trt' ;
   define nanim    / order format=8. 'Number of Animals/ Pen/ Block/ Site/ Trt';
   define tanim_site  / order format=8. 'Number of Animals/ Site/ Trt';
   define tanim    / order format=6. 'Total Number of Animals/ Trt';
   define tpen   / order format=8. 'Total Number of Pens/ Trt' ;
   define tpen_site     / order format=8. 'Number of Pens/ Site/ Trt';
   define percent1  / display format=8.1 'Conv(%)';
   define percent2  / display format=8.1 'Indep';
   define percent3  / display format=8.1 'Delta';
   define percent4  / display format=8.1 'OR';
```

```
footnote1 j=left height=8pt "Variability (Logit Scale):";
footnote2 j=left height=8pt "    Site  = &vs";
footnote3 j=left height=8pt "    Site*Trt = &vst";
footnote4 j=left height=8pt "    Block   = &vb";
footnote5 j=left height=8pt "    Pen  = &vp";
footnote6 ;
footnote7 j=left height=8pt "CRD(*)  1. Number of Animals/pen/blk/site/trt =
                                                 Number of Animals/pen/site/trt";
footnote8 j=left height=8pt "        2. Block Variability = 0";
footnote9 ;
footnote10 j=left height=8pt "RCBD(*) Number of pens/blk/site/trt = 1 by default";
run;
quit;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
<p style="margin-bottom: 3pt; margin-top: 3pt; line-height: 1;">Power calculations for complex designed clinical trials using linear mixed models

Richting: **Master of Statistics-Biostatistics**
Jaar: **2014**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Cabrieto, Jedelyn

Datum: **10/09/2014**