

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture


2008 - 20th Annual Conference Proceedings

STATISTICAL ISSUES IN EFFICACY EVALUATION FOR COMPANION ANIMAL DRUG DEVELOPMENT

Zhanglin Lin Cui

Wherly Hoffman

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

 Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Cui, Zhanglin Lin and Hoffman, Wherly (2008). "STATISTICAL ISSUES IN EFFICACY EVALUATION FOR COMPANION ANIMAL DRUG DEVELOPMENT," *Conference on Applied Statistics in Agriculture*.
<https://doi.org/10.4148/2475-7772.1099>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

STATISTICAL ISSUES IN EFFICACY EVALUATION FOR COMPANION ANIMAL DRUG DEVELOPMENT

Zhanglin Lin Cui and Wherly Hoffman
Elanco Animal Health R&D Statistics
Eli Lilly and Company
2001 W. Main St., Greenfield, IN 46140

ABSTRACT

Companion animals, commonly called pets, are animals such as dogs, cats, and horses. The companion animal drug market has expanded rapidly in recent years. Two major points of focus in companion animal drug development are therapeutics and parasiticides. From a statistics point of view, experimental design, experimental unit determination, sample size estimation and re-estimation, treatment design, data transformation, multiple testing, and proper modeling are major statistical issues when efficacy evaluation in a companion animal study is conducted. These major statistical issues are addressed using two clinical studies as examples: Reconcile® (Fluoxetine) for the treatment of separation anxiety in dogs and Comfortis® (Spinosad) for the control of fleas in dogs.

Keywords: experimental design, experimental unit, sample size determination, data transformation, multiple testing, statistical model, and companion animal.

1. INTRODUCTION

1.1. Companion Animal Drug Market

Companion animals, commonly called pets, are animals such as dogs, cats, and horses. The global companion animal health drug market generates about \$7 billion dollars of revenue each year, accounting for about 40% of the total animal health drug market. In the United States (US), pharmaceutical companies make drugs, devices, and food additives for over 100 million companion animals. The US is the dominant companion animal health drug market accounting for over 50% of the global sales. On average, a US pet owner pays as much as \$500 annually for health care for their dogs and/or cats.

1.2. Major Points of Focus in Companion Animal Drug Development

Therapeutic and parasiticide products are the major points of focus in companion animal health drug research and development. Like human beings, companion animals have many therapeutic needs. It is estimated that 17% of the dogs suffer from separation anxiety. It is estimated that 34% of all pets in the US are judged to be either overweight or obese. This mirrors the level of human obesity which has been described as “epidemic.” Pests such as fleas, lice, ticks, flies, mosquitoes, worms, and nematodes are common to dogs and cats. In the US, it is estimated that 46% of the dogs are attacked by fleas throughout the year.

1.3. Efficacy Evaluation

The objective of an efficacy evaluation is to assess the effectiveness of a new treatment for a disease or condition compared to a negative control (no treatment or placebo) or a positive control (existing treatment). In addition, the safety of the new treatment is evaluated along with cost efficiencies and ease of administration. Efficacy evaluation occurs in Phase II and Phase III clinical development. Phase II is early-stage controlled clinical studies, sometimes called proof of concept studies. They are conducted to obtain preliminary data on the effectiveness of the drug. This phase of testing also helps determine the common short-term side effects and risks associated with the drug. Phase II studies are typically well-controlled (e.g. laboratory conditions), closely monitored, and conducted using a relatively small number of companion animals, usually a few dozens of companion animals or less. Phase III studies are expanded controlled and uncontrolled trials. They are performed after preliminary evidence of effectiveness of the drug has been obtained in Phase II studies, and are intended to gather additional information on effectiveness and safety needed to evaluate the overall benefit-risk relationship of the drug. Phase III studies also provide an adequate basis for extrapolating the results to the general animal population and transmitting that information into the veterinary drug label. Phase III studies usually include several hundred companion animals in multiple veterinary clinics (commonly called sites).

This paper is an informative narrative about issues in study design and analysis in companion animal efficacy evaluation studies. Limited discussions and authors' experiences, but not detailed statistical methods on these issues, will be provided in this paper.

2. STUDY DESIGN ISSUES

To achieve the objectives of the efficacy evaluation, an efficacy study should be designed appropriately. Major statistical issues in study design include experimental unit determination, sample size estimation and re-estimation, and treatment design.

2.1. Experimental Unit in Companion Animal Health Studies

A clear definition of experimental unit establishes the foundation for the corresponding statistical analysis model. Experimental unit is defined as the smallest unit to which a treatment is randomly assigned and applied. However, it occurs in the literature that the statistical analysis method does not match the experimental unit. In companion animal studies, the experimental unit could be an individual animal, a household of one or more animals, or a pen of multiple animals.

For injection or tablet treatment, an individual animal should be the appropriate experimental unit. If a drug is given in feed or drinking water to a pen of animals as a whole, then the pen is the experimental unit although the observational unit may still be an individual animal. If a drug is applied to the animals of a house, then the household is the experimental unit even though the measurement is made on each animal in the house. When client-owned dogs or cats are enrolled in a flea clinical study, the experimental unit should be a household. All the dogs or cats in the

house should be enrolled and assigned to a same treatment group. This is because the dogs or cats in a household interact with each other, fleas can jump from one pet to another, and the responses of the treatment are highly correlated among pets within a house.

When a pen or a household is the experimental unit and individual animal is an observational unit, continuous and categorical data are handled differently. For continuous variables, the mean is initially calculated for each pen. For categorical variables (e.g. success vs. failure), a proportion is calculated for each category for each pen. When appropriate, an arcsine square root transformation is applied to the proportion data and logarithmic transformation to the count data. Then, a linear mixed model can be used to analyze the pen-based means or a generalized linear mixed model to analyze the pen based proportions. The residual variance component is used to determine sample size for future studies.

2.2. Sample Size Estimation and Re-estimation

Sample size can include the number of sites, number of blocks (or replicates), number of households (or pens), and number of animals, but not all of them in every study. The method for sample size determination depends on the statistical distribution of the primary endpoint in the study, the design of the study, and the proposed statistical analysis method. In sample size estimation, the welfare of the study animals should be considered as well. Both Good Clinical Practice (GCP) and Good Laboratory Practice (GLP) account for the welfare of the study animals ensuring, for example, that a study is not over powered.

When the primary endpoint is binary, it is assumed, in a binomial model, that each individual animal has the same underlying probability of an event (homogeneity assumption) and that outcomes are independent (independence assumption). If the observed sampling variance is higher than the variance of the theoretical model, overdispersion or extra-binomial variation occurs. Conversely, underdispersion can occur as well. This reflects a lack of independence or heterogeneity among individual animals. Extra-binomial variability can be caused by physical location of the pens, interaction or correlation among animals within a pen or household. If overdispersion or extra-binomial variation is known to occur then this should be considered in sample size estimation.

Many Phase III clinical efficacy studies are multi-site trials with a randomized complete block design (RCBD) at each site. Sample size determination for multi-site trials includes the following three steps: (1) Estimate variance components for residual (σ_e^2) and site*treatment interaction (σ_{tr}^2) from a pilot study or historical data, (2) Compute the variance of differences between treatment means ($\text{Var}(\bar{y}_{i,1} - \bar{y}_{i,2})$), and (3) Calculate sample size as the number of sites for a specified number of blocks per site. Details of the sample size determination for multi-site trials with RCBD at each site can be found in Cui et al. (2007). As an example, values of variance components and resulting sample sizes are given in Table 1. Note that for larger number of blocks per site, the number of sites required is reduced. The change in the number of sites depends on the relative size of the variance components. The sample size is calculated with a 2-sided paired t-test at the 0.05 significance level and 80% power using nQuery Advisor 7.0. (2007).

Before a clinical study is conducted, sample size can be estimated based on expected treatment differences and their standard deviations estimated from a pilot study or historical data. It is not uncommon that the initial estimation of the treatment difference and its standard deviation turns out to be too large or too small, which consequently leads to an underpowered or overpowered study. Adaptive designs which use available data from current study to modify the design as the study is in progress, as well as adjust the Type I error rate, are a possible solution to some sample size problems (Chow and Chang, 2006). For example, a pet clinical study uses available outcomes to adjust the sample size and treatment assignment of future pets, assigning more animals to the better treatment groups. Adaptive designs can offer significant ethical and cost advantages over standard fixed sampling procedure. This is called learn and confirm procedure. Adaptive designs have several undesirable properties, including lesser statistical efficiency, risks to the integrity and credibility of the trial, and overemphasis of the importance of statistical significance relative to clinical significance (Fleming, 2006).

2.3. Treatment Design

Treatment design depends on the objective of a study. The core of treatment design includes dose level determination and control selection. Treatment design determines statistical analysis methods. For example, studies with multiple dose levels often come with dose response analyses (Cui et al., 2005) as well as multiple testing. A negative (placebo) controlled study results in a superiority test while a positive (active) controlled studies result in a superiority, non-inferiority, or equivalence test.

Figure 1 illustrates the differences among the equivalence, non-inferiority, and superiority tests when a confidence interval approach is used in a positive control study. Before the study, an equivalence margin Δ should be chosen and concurred with regulatory authorities, such as Food and Drug Administration (FDA) and European Medicines Agency (EMA). Based on study data, a confidence interval of the treatment differences is computed. If the lower limit of the confidence interval is greater than zero, superiority is achieved. If the lower limit of the confidence interval is greater than minus delta, non-inferiority is achieved. If the confidence interval is between minus delta and plus delta, equivalence is achieved. If the lower limit of the confidence interval is less than minus delta, the efficacy of a new treatment can not be claimed.

3. STATISTICAL ANALYSIS ISSUES

To achieve the objectives of an efficacy evaluation, an efficacy study should be analyzed appropriately. Data transformation, multiple testing, and proper modeling are a few important issues in statistical analysis for efficacy evaluation.

3.1. Data Transformation

A data transformation should be performed when data deviates from the assumptions underlying the analysis of variance (ANOVA). The assumptions include (1) additivity: treatment and environmental effects are additive; (2) normality: experimental errors are randomly, independently, and normally distributed about zero; (3) homogeneity of variance: variances of

different treatment groups or different factor levels are homogenous. Deviation from the assumptions can affect the level of significance and the sensitivity of F or t to real departures from the null hypothesis (Steel and Torrie, 1980). Common data transformations include square root, logarithmic, and arcsin square root.

Data transformation helps but not always. If transformation does not work, alternative approaches should be taken, such as non-parametric methods (Wasserman, 2005) and weighted analyses by estimating variance for each heterogeneous group (SAS® Institute Inc., 1999).

3.2. Multiple Testing

A common problem in clinical efficacy evaluation is multiple tests of hypotheses without controlling for type I error rate. Consequently, multiplicity could result in declaring effectiveness when it is not there. Researchers could then run the risk of a wrong decision to move a compound forward in the drug development pipeline.

Multiplicity is common in clinical efficacy studies in which there are multiple dose levels, multiple treatment groups, multiple endpoints, repeated measures over time, interim analyses during the course of a clinical study, stepwise methods to find an optimal analysis, or sub-population analyses. When multiplicity happens, families of tests of hypotheses are performed in clinical efficacy studies. For example, dose response contrasts, pairwise comparisons of treatment groups, pairwise comparisons with the control or best treatment group, pairwise comparisons of repeated measures with the baseline, and pairwise comparisons of repeated measures with negative/positive/best treatment group are all cases in which multiple tests of hypotheses are performed.

Among the common P-value adjustment methods, Bonferroni and Sidak methods provide similar and the most conservative multiplicity test results, i.e. fewest significant differences. The Holm's step-down and Hochberg's step-up Bonferroni methods provide similar but less conservative results. The False Discovery Rate (FDR) method provides the least conservative results (Cui et al., 2005; Westfall et al., 1999).

3.3. Proper Modeling

Proper modeling is a big issue in a clinical efficacy evaluation. George E.P. Box once said "all models are wrong, some are useful". Although there is no perfect model, we can usually find a model that works better than others.

A big debate has been around the study site (or clinic, center). Both human and animal clinical studies end up with a pooled analysis across all study sites or centers. However, there is a big difference in the model. Site is almost always treated as random in companion animal clinical studies, but site (or center) is almost always treated as fixed in human clinical studies. Should it be random or fixed? This has to be decided during the design of the study; however a veterinary practitioner has little interest in drawing conclusions that apply only to the sites that have enrolled pets.

The following are some typical models in companion animal efficacy evaluation.

Model I: Multi-site clinical efficacy study with a completely random design (CRD) at each site

$$Y = \mu + \text{Trt} + \text{Site} + \text{Trt}*\text{Site} + \varepsilon$$

where Trt is fixed; Site is random (usually 10 – 30 sites); Trt*Site is random; and ε is the random inter-animal variability.

Model II: Multi-site clinical efficacy study with a RCBD at each site

$$Y = \mu + \text{Trt} + \text{Site} + \text{Trt}*\text{Site} + \text{Block}(\text{Site}) + \varepsilon$$

where Trt is fixed; Site is random; Trt*Site is random; Block(Site) is random; and ε is the random inter-animal variability. When a study has only a few sites, the default denominator degrees of freedom for testing treatment differences may be too small to have adequate power for the test. Alternative approaches should be considered.

Model III: Multi-site clinical efficacy study with CRD at each site and repeated measures

$$Y = \mu + \text{Trt} + \text{Site} + \text{Trt}*\text{Site} + \varepsilon + \text{Time} + \text{Trt}*\text{Time} + \text{Site}*\text{Time} + \delta$$

where Trt is fixed; Site is random; Trt*Site is random; ε is random inter-animal variability; Time is fixed; and δ is random intra-animal variability.

Model IV: Multi-site clinical efficacy study with household type as a fixed effect and repeated measures

$$Y = \mu + \text{Site} + \text{House} + \text{Trt} + \text{Trt}*\text{Site} + \text{Trt}*\text{House} + \varepsilon + \eta + \text{Time} + \text{Time}*\text{Site} + \text{Time}*\text{House} + \text{Time}*\text{Trt} + \text{Time}*\text{Trt}*\text{House} + \delta + \zeta$$

where Site is random; House is fixed, representing household type (e.g. single pet household vs. multiple pets household); Trt is fixed; Time is fixed; ε is random variability among households; η is random variability among dogs (within households); δ is random intra-household variability; and ζ is random intra-animal variability. In this model, the household is the experimental unit and individual animal is the experimental sub-unit (Keefe, 2007).

A statistical model does not have to be complicated. A complicated model may cause convergence problem, may be hard to explain the effects, and sometimes may not have appropriate error terms for testing. Statisticians care about the “true” treatments difference and p-values. Pet owners care about the gross treatment success for their dogs or cats. Veterinary practitioners care about practically meaningful treatment differences. A statistically significant difference may be meaningless to a veterinary practitioner. Conversely, a numerical treatment difference may be meaningful to a veterinarian although it is not statistically significant. This occurs when sample size is too small to detect the effect.

4. EXAMPLES

4.1. Reconcile for Treatment of Separation Anxiety in Dogs

Separation anxiety (SA) results in problematic behaviors when a dog is left alone, including destructive/rearranging behavior, excessive vocalization, inappropriate urination, inappropriate defecation, excessive salivation, excessive licking or grooming, shaking or shivering, restlessness, and depression. Separation anxiety is common in some dogs. It is a treatable condition.

A multi-centered, double-masked, controlled, parallel-arm clinical Phase III study was conducted in 42 animal clinics (sites) in the US and Canada. The primary objective was to confirm the clinical effectiveness of 1-2 mg/kg of Reconcile chewable tablets administered once daily for the treatment of SA in client-owned dogs.

Each individual dog was an experimental unit. Power calculations were based on a generalized linear mixed model for logistic regression. To account for Site*Treatment interaction, an overdispersion factor of 10% was applied to the binomial variation. Based on 100 dogs per treatment group, this study had approximately 80% power to detect a difference of 25% between Reconcile and placebo in the incidence of improved overall SA score. Two hundred and twenty nine (229) healthy, client-owned dogs were enrolled in the study. Of those 229, 112 dogs received the control tablets (placebo) and 117 received Reconcile chewable tablets. All dogs, regardless of which arm, received behavior modification performed by owners. Behavior modification should help both arms. Ninety-two dogs in the control group and 96 dogs in the Reconcile chewable tablets group completed the study to be evaluated for effectiveness.

The primary effectiveness measure was the incidence of dogs demonstrating improvement in overall SA severity score (owner's impression of signs of SA) in the 8-week treatment period compared with the 2-week pretreatment baseline. Incidence of improvement of the individual SA behaviors and frequency of test article palatability were also assessed. Individual SA behaviors, such as inappropriate urination, inappropriate defecation, and others, were also evaluated.

The statistical significance of treatment on the improvement on the owner assessed behavioral severity scores was assessed using a generalized linear mixed model analysis with the logit link function and the binomial error distribution. The effect of treatment upon the binary response variable, improved or not improved relative to baseline, was modeled separately for each week. In the analysis, study center and study center by treatment interactions were treated as random effects.

The percentage of dogs with improved global SA scores was statistically higher for dogs that received Reconcile chewable tablets compared with dogs who received control tablets in each week, with the exception of treatment week 3 (Figure 2).

4.2. Comfortis for Control of Fleas on Dogs

Fleas are found in every region of the United States. They can be a major problem for dogs and dog owners. A dog may be exposed to these blood-sucking parasites anywhere: backyard, public parks, beaches, playgrounds, wooded areas, even at home. Each female flea can produce 40 to 50 eggs a day, up to 2,000 in their lifetime. Even a few fleas can turn into a major infestation in a matter of days.

A masked, parallel arm, controlled, laboratory study was conducted in North Carolina (NC) and California (CA). The objective was to confirm the 30-day residual effectiveness of Comfortis tablets (dosed once orally at a dose range of 30-40 mg/kg) for the treatment of adult cat flea, *Ctenocephalides felis*, infestations on dogs under laboratory conditions.

The individual dog was the experimental unit. The sample size was determined based on the statistical power to show significance according to FDA guidelines for the evaluation of flea efficacy. Each study site was powered individually for single site efficacy evaluation. In each site, 24 dogs were included in a randomized complete block design with dog gender and pretreatment live flea counts as blocking factors.

Each dog was infested with approximately 100 newly emerged, unfed adult fleas on test days -1, 5, 12, 19, 28, and 35. Dogs were dosed once on test day 0 and were fed at the time of dosing. Adult flea counts, at approximately 48 hours post-treatment or post-infestation, were conducted on test days 2, 7, 14, 21, 30, and 37 to assess effectiveness. Percent effectiveness was based on geometric means of live flea counts on Comfortis treated and control dogs on each day of assessment.

Neither of the two transformations (Logarithm, Square root) resulted in normally distributed residuals because most of the observations of the Comfortis treated group were basically zeros. Therefore, a logarithmic transformation was selected because it is more commonly used for count data than a square root transformation in clinical studies. More specifically, it is typically used in flea clinical studies.

The log-transformed data were analyzed using a linear mixed model:

$$Y = \mu + \text{TRT} + \text{Gender} + \text{TRT} * \text{Gender} + \text{Day} + \text{TRT} * \text{Day} + \text{Gender} * \text{Day} \\ + \text{TRT} * \text{Gender} * \text{Day} + \text{Block}(\text{Gender}) + \text{TRT} * \text{Block}(\text{Gender}) + \varepsilon$$

Since the logarithmic transformation did not resolve the normality problem satisfactorily, non-parametric tests were performed for treatment effects at each time point using PROC NPAR1WAY (SAS® Institute Inc., 1999).

Percent efficacy was calculated based on the geometric means for the treatment groups:

$$\% \text{Efficacy} = \frac{\text{Geometric Mean}_{\text{Control}} - \text{Geometric Mean}_{\text{Drug}}}{\text{Geometric Mean}_{\text{Control}}}$$

The geometric means of live flea counts were different ($P < 0.001$) between treatment groups on each test day (Table 2). Adjusted p-values using Hochberg's Step-up Bonferroni methods ended

up with the same p-values as in Table 2. The linear mixed model analysis produced the same conclusion as the non-parametric tests. Post-treatment geometric means of live flea counts ranged from 69.4 to 82.3 in the control group and 0 to 1.9 in the Comfortis treated group, respectively. The percent effectiveness ranged from 96.4 to 100% through test day 30 (Table 2).

5. SUMMARY AND FUTURE PROSPECTS

In summary, experimental design, experimental unit determination, sample size estimation and re-estimation, and treatment design are important issues in study design in companion animal drug development; while data transformation, multiple testing, and proper modeling are major issues in statistical analysis. Looking into the future, adaptive design and modeling with pen or household as experimental unit are potentially beneficial for companion animal drug efficacy evaluation.

6. ACKNOWLEDGEMENTS

The authors would like to express appreciations to the reviewers of this paper whose suggestions helped improve this work. In particular, the authors wish to thank Drs. Daniel Mowrey and Alan Zimmermann of Eli Lilly and Company.

7. BIBLIOGRAPHY

Chow, S.C. and M. Chang. 2006. Adaptive Design Methods in Clinical Trials. CRC Press/Chapman Hall/Dekker, Taylor & Francis, New York, New York.

Cui, Zhanglin, Alan Zimmermann, Daniel H. Mowrey. 2007. Sample Size Determination in Animal Health Studies. *In Proceedings of Applied Statistics in Agriculture 2007*, pp143-153, Kansas State University.

Cui, Zhanglin, Daniel H. Mowrey, Alan G. Zimmermann, and Douglas E. Hutchens. 2006. A Comparison of Multiple Tests Procedures: Spinosad as a Treatment for Lice on Cattle. *In Proceedings of Applied Statistics in Agriculture 2006*, pp5-17, Kansas State University.

Cui, Zhanglin, Daniel H. Mowrey, Alan G. Zimmermann, James T. Symanowski, Howard B. Green, and John I. D. Wilkinson. 2005. The Effect of Monensin on Lactation Dairy Cows: A Dose Response Evaluation. *In Proceedings of Applied Statistics in Agriculture 2005*, pp102-111, Kansas State University.

Fleming, Thomas R. 2006. Standard versus adaptive monitoring procedures: A commentary. *Statistics in Medicine* 25:3305-3312.

Keefe, Thomas J. 2007. Issues in the Design and Analysis of Clinical Studies of Companion Animals. FDA/Industry Workshop in Statistics, Washington, D.C., September 18, 2007.

nQuery Advisor 7.0. 2007. Statistical Solutions, Stonehill Corporate Center, Suite 104, 999 Broadway, Saugus, MA 01906, USA.

SAS Institute Inc. 1999. SAS/STAT User's Guide, Version 8, Volume 2. Cary, NC: SAS Institute Inc.

Steel, R.G. and J.H. Torrie. 1980. Principles and procedures of statistics, A biometrical approach. McGraw-Hill, Inc., New York, 2nd edition.

Wasserman, Larry. 2005. All of Nonparametric Statistics. Springer. New York.

Westfall, P.H., R.D. Tobias, D. Rom, R.D. Wolfinger, and Y. Hochberg. 1999. Multiple comparisons and multiple tests using the SAS system. SAS Institute Inc., Cary, NC, USA.

**Table 1. Sample Size Determination for Multi-Site Studies
 Using a Randomized Complete Block Design**

Site*Treatment Variance Component (s^2_{tr})	Residual [Treatment*Block(Site)] Variance Component (s^2_e)	Number of Blocks per Site (J)	Variance of Treatment Differences $2*(s^2_{tr} + s^2_e/J)$	Standard Deviation of Treatment Differences	Difference to Detect	Sample Size (Number of Sites Required)
20	12	2	2*26	7.211	6	14
20	12	3	2*24	6.928	6	13
20	12	4	2*23	6.782	6	13
20	12	5	2*22.4	6.693	6	12
10	12	2	2*16	5.657	6	10
10	12	3	2*14	5.292	6	9
10	12	4	2*13	5.099	6	8
10	12	5	2*12.4	4.980	6	8

**Table 2. Percent Efficacy and P-values in the Study on Comfortis
 for Control of Fleas on Dogs**

Percent efficacy (%)						
Study Site	Day 2	Day 7	Day 14	Day 21	Day 30	Day 37
California	100	100	100	100	99.9	97.7
North Carolina	99.9	100	99.6	99.3	96.4	87.4
P-values [1]						
Study Site	Day 2	Day 7	Day 14	Day 21	Day 30	Day 37
California	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
North Carolina	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

[1] Non-parametric Wilcoxon exact test

Figure 1. Equivalence, Non-inferiority, and Superiority Tests Using Confidence Interval Approach

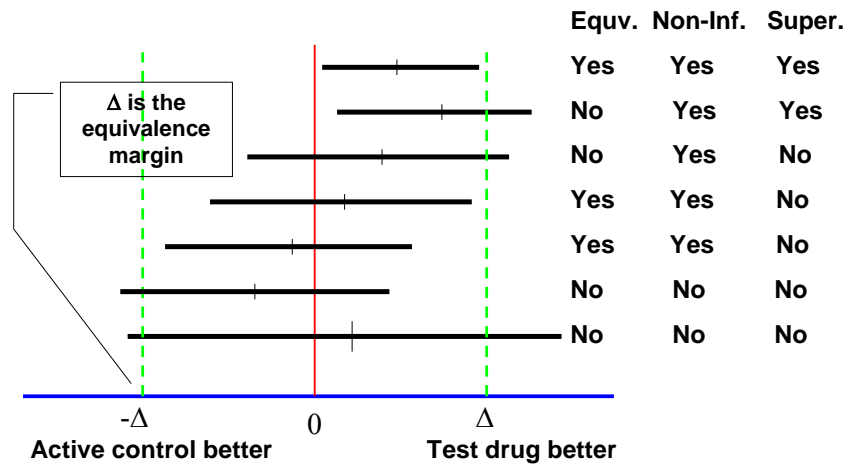


Figure 2. Statistical Significance at Each Treatment Week

