

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2007 - 19th Annual Conference Proceedings


SAMPLE SIZE DETERMINATION IN ANIMAL HEALTH STUDIES

Zhanglin Cui

Alan G. Zimmermann

Daniel H. Mowrey

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; Zimmermann, Alan G.; and Mowrey, Daniel H. (2007). "SAMPLE SIZE DETERMINATION IN ANIMAL HEALTH STUDIES," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1114>

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.

SAMPLE SIZE DETERMINATION IN ANIMAL HEALTH STUDIES

Zhanglin Cui, Alan G. Zimmermann and Daniel H. Mowrey
Statistics and Information Sciences
Eli Lilly and Company
2001 W. Main St., Greenfield, IN 46140

ABSTRACT

Oftentimes in animal health studies, a treatment group is randomly assigned to a pen of animals, and the pen of animals as a whole is treated (fed the same medicated feed or water) together. In this scenario, the pen of animals is the experimental unit and the individual animal may be an observational unit. In addition to having the pen as the experimental unit, if multiple sites are used and site is treated as a random factor, this adds complexity to the study. To properly design the study, it is necessary to determine the number of animals in a pen, the number of pens per treatment group, and the number of sites in order to detect treatment differences with desired power. 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. Focusing on the case where individual animal is not the experimental unit, this paper demonstrates methods and discusses issues for sample size determination in animal health studies with continuous or binary primary endpoints, and gives a simple approach for determining an appropriate number of sites in a multi-site model analyzed with a linear mixed model.

Key words sample size, power, experimental unit, experimental design, randomized complete block design, binomial distribution, normal distribution

1. INTRODUCTION

Sample size is important in making statistical inferences. A researcher should use an adequate sample size to effectively address research questions. Sample size usually refers to the number of experimental units in a treatment group. The experimental unit is defined as the smallest unit to which a treatment is randomly assigned and applied.

This paper demonstrates methods and discusses issues for sample size determination in animal health studies with continuous or binary primary end points.

2. EXPERIMENTAL UNIT IN ANIMAL HEALTH STUDIES

In animal health studies, the experimental unit could be an individual animal, a pen, a tank, a raceway, or a herd. A treatment group is randomly assigned to a pen of animals, and the pen of animals as a whole is treated (given the same medicated feed or water) together. In this scenario,

the pen of animals is the experimental unit and the individual animal may be an observational unit. In addition to having the pen as the experimental unit, multiple trial sites treated as a random factor create additional complexity for the study design.

To properly design the study, it is first necessary to estimate the pen-to-pen variability and determine the number of animals in a pen (i.e. pen size), then the number of pens per treatment group (i.e. sample size), and finally, for multi-site studies, the number of sites necessary to detect treatment differences at a predetermined power. For multi-site studies with a randomized complete block design (RCBD) within each site, the sample size is the product of the number of blocks per site and the number of sites.

3. ESTIMATION OF PEN-TO-PEN VARIABILITY

When pen is the experimental unit and also the observational unit in a completely randomized design, the observation y_{ij} on i^{th} treatment and j^{th} pen can be fitted in a linear model,

$$y_{ij} = \mu + \tau_i + e_{ij} \quad (1)$$

where μ is the overall mean, τ_i the fixed treatment effect, and e_{ij} the random pen effect with $e_{ij} \sim N(0, \sigma_{\text{pen}}^2)$ where σ_{pen}^2 is the pen-to-pen variability. Often, the pen-to-pen variability can be estimated from a set of historical data or pilot studies.

When pen is the experimental unit and individual animal is the observational unit in a completely randomized design, the above model is still the most appropriate model. To apply the model, the means for a continuous endpoint, the means for a count endpoint, or the proportions for a binary endpoint should be calculated for each pen. Appropriate data transformations may be needed prior to model estimation to meet the normal distribution assumption. The variance components for residual, e_{ij} , can be then calculated from the pen-based mean model and used in sample size determinations for future studies.

To further examine the variance components for the pen-to-pen variability when pen is the experimental unit and individual animal is the observational unit, the observation y_{ijk} on i^{th} treatment, j^{th} pen, and k^{th} animal can be modeled as,

$$y_{ijk} = \mu + \tau_i + p_{ij} + e^*_{ijk} \quad (2)$$

where μ is the overall mean and τ_i is the fixed treatment effect. The random pen effect $p_{ij} \sim N(0, \sigma_p^2)$, σ_p^2 is the variance component due to pen while $e^*_{ijk} \sim N(0, \sigma_e^2)$, σ_e^2 is the variance component due to animal.

Averaging equation (2) over animals within pens yields:

$$\bar{y}_{ij\bullet} = \mu + \tau_i + p_{ij} + \bar{e}_{ij\bullet}^* \quad (3)$$

The sum of p_{ij} and $\bar{e}_{ij\bullet}^*$ is equivalent to e_{ij} of pen as observational unit. The pen-to-pen variability, $\sigma_{\text{pen}}^2 = \text{Var}(p_{ij} + \bar{e}_{ij\bullet}^*) = \sigma_p^2 + \sigma_e^2/m$, which includes variability due to m animals (σ_e^2/m) and variability due to pen (σ_p^2). Increasing the number of animals per pen will reduce part of the pen-to-pen variability. However, it is not possible to separate these two components from the model $y_{ij} = \mu + \tau_i + e_{ij}$ which is more appropriate when pen is the experimental unit and individual animal is the observational unit in a completely randomized design. In addition, the typical assumption of independence among the e_{ijk}^* (i.e. independence among animals in a pen) is not true in practice and the variability due to animal may not be exactly σ_e^2/m .

In multi-site studies, the variance components for both the residual and the site*treatment interaction terms must be calculated in order to determine the number of sites and number of replications. For binary variables, pen-to-pen variability can be calculated based on the binomial distribution if there is no extra-binomial variation. More details on the pen-to-pen variability for the binary variables will be presented in later sections.

4. PEN SIZE DETERMINATION

Ideally, the most appropriate way to determine the pen size is to conduct a study with a varying number of animals per pen. The variability among pens could then be calculated for each of the different pen sizes with the general expectation that the variability decreases when the pen size increases. Studies such as this are useful to determine pen size; however they do not determine sample sizes with regard to number of experimental units (pens). If pen size cannot be determined by experimentation, then a subjective method is required.

To determine the pen size subjectively, many different practical factors may need to be considered, such as commercial feedlot practice, regional differences, animal management standards, and budget. Sometimes, animal biological characteristics and limitations also play important roles in pen size determination. For example, in pig studies, pen size should be less than 20 head per pen. This is because pigs can only recognize and live harmoniously with a limited number of pen mates. Pens with more than 20 animals usually result in fighting among the pigs before the establishment of a harmonious social order and a study with many injured and early removed pigs. In other cases with mixed genders, intact males and non-pregnant females are required and pen size can be as small as one animal per pen, or more than one animal per pen if genders are separated in different pens or if castrated males and/or spayed females are used.

As shown earlier, pen size determines part of the pen-to-pen variability and, in turn, sample size. If m is the pen size, i.e. the number of animals per pen, the pen-to-pen variability decreases as m increases. Assuming there is no physical pen effect and responses of animals within a pen are

independent, the observed pen-to-pen variability s_{pen}^2 has an expected value $\sigma_{\text{animal}}^2/m$. In fact, the above assumption is not true in animal health studies. Animals within a pen interact with each other, creating additional covariances. There may be other differences between pens that are not directly related to how many animals are in the pen as well. The animals are of course different in each pen, creating a unique pen effect. In brief, there is pen-to-pen variability apart from what the animal variability with variance that changes proportionally with $1/m$.

If the endpoint is binary in nature, e.g. success versus failure, then Figure 1 gives some guidance on the number of animals per pen. It can be seen that, as number of animals per pen increases, the pen-to-pen standard deviation (SD) decreases, resulting in a decrease in sample size (number of pens) and increase in total number of animals. The total number of animals per treatment group is the product of number of animals per pen and the sample size. More details about Figure 1 will be given in Section 5.2.

5. SAMPLE SIZE DETERMINATION

5.1. Continuous Endpoint

Continuous endpoints in animal health studies, such as body weight, milk yield, lean meat yield, and protein content, measure quantitative responses of animals to a treatment. Assume a comparative study with two treatment groups, n experimental units per treatment group, and m animals per pen. Here, the pen is the experimental unit while individual animals are the observational units. Assume that a completely randomized design is used to assign treatments to pens.

Let \bar{Y}_A and \bar{Y}_B be the pen means of a continuous normally-distributed variable for treatment groups A and B, respectively. Assume that both groups have a common variance σ^2 with estimate s^2 from a linear mixed model or ANOVA based on $2n-2$ degrees of freedom. The mean treatment difference follows a normal distribution,

$$(\bar{Y}_A - \bar{Y}_B) \sim N\left(\mu_A - \mu_B, \frac{2\sigma^2}{n}\right)$$

which may be simplified as $D \sim N\left(\delta, \frac{2\sigma^2}{n}\right)$,

where D is the observed treatment differences and δ is the expected mean difference. The quantity $\frac{D - \delta}{\sigma\sqrt{\frac{2}{n}}}$ follows the standard normal distribution,

$$\frac{D - \delta}{\sigma\sqrt{\frac{2}{n}}} \sim N(0, 1)$$

If the test size is α , the one-sided critical value will be $Z_\alpha \sigma \sqrt{\frac{2}{n}}$. Denote the power to be $1-\beta$, thus the probability to have the observed treatment difference exceed the critical value is given by:

$$P(D > Z_\alpha \sigma \sqrt{\frac{2}{n}}) = 1 - \beta$$

Now setting the quantity $\frac{D - \delta}{\sigma \sqrt{\frac{2}{n}}}$ equal to $-Z_\beta$, then

$$\frac{D - \delta}{\sigma \sqrt{\frac{2}{n}}} = -Z_\beta \quad \text{or} \quad D = \delta - Z_\beta \sigma \sqrt{\frac{2}{n}}$$

To ensure the power, the observed treatment difference D must exceed the critical value $Z_\alpha \sigma \sqrt{\frac{2}{n}}$, i.e.,

$$\delta - Z_\beta \sigma \sqrt{\frac{2}{n}} \geq Z_\alpha \sigma \sqrt{\frac{2}{n}}$$

Solving for n (per treatment group), we obtain

$$n \geq \frac{2(Z_\alpha + Z_\beta)^2 \sigma^2}{\delta^2} \quad \text{for one-sided test, and}$$

$$n \geq \frac{2(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{\delta^2} \quad \text{for two-sided test,}$$

where σ^2 is the pen-to-pen variability estimated by the observed s^2 ; δ is the expected mean treatment difference estimated by the observed treatment difference based on the pen means.

In general, the student t-distribution is often used instead of normal distribution to calculate the sample size leading to larger estimates.

Example I. Compound R in feed for growth performance in turkeys. In this study, the primary efficacy variable was average daily gain (ADG, kg/day/head). A pen size of $m = 10$ turkeys per pen was determined by the investigator. Based on a previous pilot study, the pen-based ADG for the control group was 0.321 kg/day/head which was assigned to be the population mean μ_B . A 7% increase in ADG is expected for the group treated with compound R in a new study. Thus the expected ADG for the treatment R group is assumed to be $\mu_A = 0.312 + 0.312 * 0.07 = 0.34347$ kg/day/head. The treatment difference is $\delta = 0.34347 - 0.321$

=0.02247. The common standard deviation derived from the residual error term of a linear mixed model analysis of the pilot study was $s = 0.0214$. Given the test size $\alpha = 0.05$, power $(1-\beta) = 80\%$, and two-sided t-test, the sample size of 16 pens per treatment group was calculated.

$$n \geq \frac{2(t_{\alpha/2, 2n-2} + t_{\beta, 2n-2})^2 s^2}{\delta^2} = 16 \text{ pens per treatment group}$$

5.2. Binary Endpoint

Binary endpoints in animal health studies, such as success versus failure, positive versus negative, present versus absent, measure qualitative responses of animals to a treatment. Assume a comparative study with 2 treatment groups, n experimental units per treatment group, and m animals per pen. Here, the pen is the experimental unit while individual animals are the observational units. Further assume that a completely randomized design is used to assign treatments to pens. The endpoint variable is treatment success or failure in a therapeutic study. The proportion of success is calculated on each pen. The drug treated group (A) has a mean success rate $\mu_A = p_A$ and variance $\sigma_A^2 = p_A(1-p_A)/m_A$; while the control group (B) has a mean success rate $\mu_B = p_B$ and variance $\sigma_B^2 = p_B(1-p_B)/m_B$. The common pen-to-pen variability can be estimated as follows.

$$s^2 \approx (\sigma_A^2 + \sigma_B^2)/2 = (p_A(1-p_A)/m_A + p_B(1-p_B)/m_B)/2$$

$$s \approx \sqrt{(p_A(1-p_A)/m_A + p_B(1-p_B)/m_B)/2}$$

Note that this estimate assumes that the pen-to-pen variability is due only to the number of animals per pen, and not due to any other differences between pens (such as physical location of pens) that may lead to extra-binomial variation. In situations where extra-binomial variation is anticipated, it is advisable to obtain pilot data or use data from past similar studies and estimate the common pen-to-pen variability from a linear mixed model analysis or ANOVA.

Example II. Compound T in feed for the control of disease B in cattle. The primary efficacy variable was the binary treatment outcome, success versus failure, for each individual animal at the end of the treatment phase (Study Day 28). The success rate was calculated as the ratio of the number of success animals to the total number of animals evaluated in a pen. In a previous study, the success rate for the compound T treated group was $p_A = 0.554$ while the success rate for the control group was $p_B = 0.314$. The treatment difference in success rates was therefore $\delta = 0.554 - 0.314 = 0.24$. To aid the design of the study, pen sizes between 1 and 30 animals per pen were considered in the sample size determination process (Figure 1). At each pen size the corresponding pen-to-pen standard deviation (SD), the sample sizes, and total number of animals were calculated based on the 80% power and a one-sided t-test at the 5% significance level. It can be seen that, as the pen size increases from 1 to 10 animals per pen, the sample size decreases dramatically; whereas, when the pen size is over 10, the sample size does not change

much. Adding additional animals beyond 10 to the pen will increase the total number of animals in the study and, in turn, increase the cost of the study without a corresponding gain in the study power. A pen size of $m = 12$ calves per pen was determined by the investigator. The estimated common standard deviation in terms of rate of success animals was calculated:

$$\begin{aligned} s &= \sqrt{(p_A(1-p_A)/m_A + p_B(1-p_B)/m_B)/2} \\ &= \sqrt{(0.554*(1-0.554)/12 + 0.314*(1-0.314)/12)/2} \\ &= 0.13882 \end{aligned}$$

Given the test size $\alpha = 0.05$, power $(1-\beta) = 90\%$, and a one-sided t-test, the sample size of 7 pens per treatment group was calculated:

$$n \geq \frac{2(t_{\alpha, 2n-2} + t_{\beta, 2n-2})^2 s^2}{\delta^2} = 7 \text{ pens per treatment group}$$

Alternatively, the common standard deviation in terms of the number of success animals could have been calculated:

$$\begin{aligned} s &= \sqrt{(m_A p_A (1-p_A) + m_B p_B (1-p_B))/2} \\ &= \sqrt{(12 * 0.554 * (1-0.554) + 12 * 0.314 * (1-0.314))/2} \\ &= 1.6658 \end{aligned}$$

The estimation of expected average number of success animals in T treated group would have been $\mu_A = 12 * 55.4\% = 6.648$ animals/pen and in the control groups $\mu_B = 12 * 31.4\% = 3.768$ animals/pen. The treatment difference in success rates would be $\delta = 6.648 - 3.768 = 2.88$. Given the test size $\alpha = 0.05$, power $(1-\beta) = 90\%$, and a one-sided t-test, the sample size of 7 pens per treatment group was again calculated:

$$n \geq \frac{2(t_{\alpha, 2n-2} + t_{\beta, 2n-2})^2 s^2}{\delta^2} = 7 \text{ pens per treatment group}$$

5.3. Continuous Endpoint with Multiple Sites

Typically, animal health phase III clinical studies are conducted at multiple sites with a randomized complete block design at each site. Common blocking factors are pen location within a facility, gender, parity, initial body weight, or initial flea count. The sample size per treatment is the product of number of blocks per site and the number of sites. Appropriate combinations of number of blocks per site and the number of sites could significantly reduce the study cost.

The study data are usually analyzed using a linear mixed model with treatment as fixed effect and site, site*treatment, and block(site) as random effects. A covariate such as initial body weight, initial flea count, or initial milk yield is often included in the model, but this is not discussed here. Site heterogeneity may be tested using Levene's test.

Briefly, sample size determination for multiple site trials include the following three steps: (1) Estimate variance components for residual (σ_e^2) and site*treatment (σ_{τ}^2), (2) Compute the variance of differences between treatment means ($\text{Var}(\bar{y}_{i\cdot 1} - \bar{y}_{i\cdot 2})$), and (3) Calculate sample size as the number of sites for a specified number of blocks per site. Details of the sample size determination follow.

The model can be expressed as

$$y_{ijk} = \mu + t_i + b_{ij} + \tau_k + (\tau\tau)_{ik} + e_{ijk}$$

where

- μ = overall mean;
- t_i = random effect of i^{th} site, $i = 1, 2, \dots, I$;
- b_{ij} = random effect of j^{th} block at i^{th} site, $j = 1, \dots, J$;
- τ_k = fixed effect of k^{th} treatment, $k = 1, \dots, K$ (often, $K=2$);
- $(\tau\tau)_{ik}$ = random effect of site*treatment interaction;
- e_{ijk} = residual.

All the random effects are treated as being independent. The variance components in the model are

$$\begin{aligned} \text{Var}(t_i) &= \sigma_t^2 & \text{Var}(b_{ij}) &= \sigma_b^2 \\ \text{Var}[(\tau\tau)_{ik}] &= \sigma_{\tau\tau}^2 & \text{Var}(e_{ijk}) &= \sigma_e^2 \end{aligned}$$

The important variance components will be those for residual (σ_e^2) and site*treatment (σ_{τ}^2). These are the only two estimates which we will need for the sample size determination.

Using a "dot" and bar to represent averaging for site*treatment means, we have

$$\begin{aligned} \bar{y}_{i\cdot k} &= \mu + t_i + \bar{b}_{i\cdot} + \tau_k + (\tau\tau)_{ik} + \bar{e}_{i\cdot k} \\ &= \mu + [t_i + \bar{b}_{i\cdot}] + \tau_k + [(\tau\tau)_{ik} + \bar{e}_{i\cdot k}] \\ &= \mu + t_i^* + \tau_k + e_{ik}^* \end{aligned}$$

where $e_{ik}^* = (\tau\tau)_{ik} + \bar{e}_{i\cdot k}$ and $\text{Var}(e_{ik}^*) = \sigma_{\tau\tau}^2 + \frac{\sigma_e^2}{J}$

and t_i^* defined similarly.

The model for difference between site*treatment means becomes a very simple one-sample model

$$\bar{y}_{i\bullet 1} - \bar{y}_{i\bullet 2} = (\tau_1 - \tau_2) + (\bar{e}_{i\bullet 1}^* - \bar{e}_{i\bullet 2}^*), \text{ for } i = 1, \dots, I$$

$$\text{Var}(\bar{y}_{i\bullet 1} - \bar{y}_{i\bullet 2}) = \text{Var}(\bar{e}_{i\bullet 1}^* - \bar{e}_{i\bullet 2}^*) = 2 \times \left(\sigma_{tr}^2 + \frac{\sigma_e^2}{J} \right)$$

With estimates for residual (σ_e^2) and site*treatment (σ_{tr}^2), for several different values of J we calculate the necessary sample size in terms of the number of sites I. Given these variances, the sample size can be calculated using software or formulas for one-sample model or a paired t-test.

Example III. Sample Size Determination for Multi-Site Studies. 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 amount of change in required number of sites will depend on the relative size of the variance components. The sample size is calculated with 2-sided test at 0.05 significance level and 80% power.

6. SUMMARY

Sample size determination includes the number of animals per pen, the number of pens per treatment group, and the number of trial sites, when designing an animal health study with the pen of animals as the experimental unit. There are processes to determine an appropriate pen size, however pen size determination is often subjective depending on multiple practical and biological factors. Pen size can affect the pen-to-pen variability and hence the number of pens, and thus the number of animals in a study. For continuous endpoints or binary endpoints, the pen-to-pen variability can be estimated from a linear mixed model analysis of pen-based means. For multiple sites studies, sample size can be calculated based on the variance components for site*treatment and residual.

7. BIBLIOGRAPHY

- Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Academic Press, New York, 2nd edition.
- Desu, M. M. and D. Raghavarao. 1990. Sample size methodology. Academic Press, Boston.
- Freiman, J.A., T.C. Chalmers, H. Smith, Jr., and R.R Kuebler. 1986. The importance of beta, the type II error, and sample size in the design and interpretation of the randomized controlled trial: Survey of 71 “Negative” Trials. *In* Medical Uses of Statistics, eds. J. C. Bailar III and F. Mosteller, chap. 14, pp. 289–304. NEJM Books, Waltham, Mass.
- Kraemer, H. C. and S. Thiemann. 1987. How many subjects? Statistical power analysis in research. Sage Publications, Newbury Park, CA.

Lipsey, M. W. 1990. Design sensitivity: Statistical power for experimental research, Sage Publications, Newbury Park, CA.

Mace, A. E. 1964. Sample-size determination. Reinhold, New York.

nQuery Advisor 6.0. 2005. Statistical Solutions, Stonehill Corporate Center, Suite 104, 999 Broadway, Saugus, MA 01906, USA.

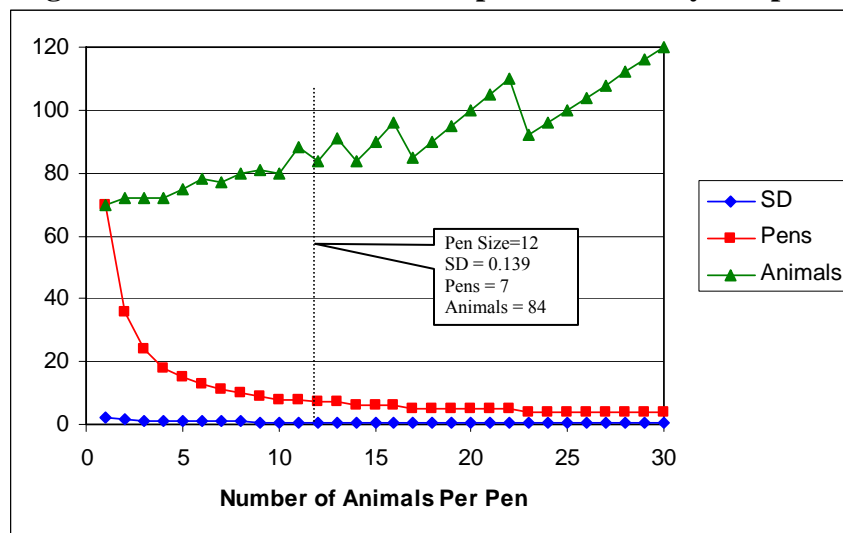
Odeh, R. E. and M. Fox. 1991. Sample size choice: Charts for experiments with linear models. Marcel Dekker, New York, second edition.

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

Table 1. Sample Size Determination for Multi-Site Studies

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 $2*(s^2_{tr} + s^2_e/J)$	Difference to Detect	Sample Size (Number of Sites Required)
12	10	2	$2*17$	20	23
12	10	3	$2*15.3$	20	19
12	10	4	$2*14.5$	20	17
12	10	5	$2*14$	20	16
5	10	2	$2*10$	20	8
5	10	3	$2*8.3$	20	6
5	10	4	$2*7.5$	20	5
5	10	5	$2*7$	20	4

Figure 1. Pen Size Effects on Sample Size – Binary Endpoint.



Note: The X-axis is the pen size (i.e. the number of animals per pen). The Y-axis is simply the magnitude of the pen-to-pen variability (i.e. standard deviation, SD), the sample size (i.e. number of pens, Pens), and the number of animals per treatment group (Animals).