

Ruminant Nutrition 6: Assessment of Feeding Management Practices for Beef and Dairy Cattle

Interpretation and Design of Nonregulatory On-Farm Feeding Trials¹

N. R. ST-PIERRE* and L. R. JONES†

*Department of Animal Sciences,
The Ohio State University,
Columbus, OH 43210

†F.A.R.M.E. Institute, Inc.,
Homer, NY 13077

ABSTRACT

Nonregulatory feeding trials are used to determine the frequency and magnitude of response to a nutritional treatment, the odds of success, and factors that influence the odds and to provide an estimate of potential economic impact. These types of trials are not appropriate for elucidating mechanisms or modes of action. Correct experimental design is critical. Important factors that impact the validity of these trials are animal randomization, potential confounding with time, identification of the proper experimental unit, and adequate replication. Animals should be assigned to treatment without a systematic influence of environmental or physiological factors. The experimental design must account for and remove these influences. The experimental unit is the smallest entity to which the treatment can be applied randomly. In commercial settings, this is often a pen of animals. In a pen feeding situation, the experimental unit is the pen, although measurements may be taken on an individual animal basis. Frequency of data collection generally does not influence the treatment effect; however, it does influence the variance associated with the observations and the power to detect differences in treatments. Statistical methods that account for repeated observations over time are required when analyzing data with multiple observations on the same animal. Improper accounting for environmental changes over time is the most common error. Other errors include improper assignment of animals to treatment, failure to replicate the treatment across multiple experimental units, lack of on-farm oversight, and poor calibration of test equipment.

¹Salaries and research support were provided by state and federal funds appropriated to the Ohio Agricultural Research and Development Center, The Ohio State University. Correspondence should be addressed to 221 Animal Science Building, 2029 Fyffe Road.

©1999 American Society of Animal Science
and American Dairy Science Association. All rights reserved.

(Key words: statistical design, feeding trials, statistical analysis)

Abbreviation key: EMS = expected mean squares, MST = multi-site trials, SST = single-site trials.

INTRODUCTION

Feed and pharmaceutical industries rely heavily on on-farm trials for evaluating feed ingredients, feed additives, and other commercial products. The legal and scientific implications of such trials are clearly of a much lesser degree than those of trials performed for regulatory purposes, like those used for a New Animal Drug Application with the Food and Drug Administration. Most would agree that the trail of documentation can be relaxed for nonregulatory on-farm feeding trials. However, this relaxation does not justify the use of improper experimental designs or analyses that lead to incorrect conclusions. The objectives of this paper are to review the proper planning, design, and analyses of on-farm feeding trials, and to highlight common mistakes that are frequently made.

ON-FARM FEEDING TRIALS

Numerous reasons justify on-farm feeding trials. First, such trials can determine the distribution (mean and dispersion) of response to a new feeding technology. Second, they are used to estimate the frequency of success under a wide range of conditions and to estimate management and environmental factors that impact success. Third, their results can be used to determine potential economic impacts of a new practice and serve as a basis for decision-making (4, 8). Fourth, they offer a cost-efficient alternative to research herds when a large number of experimental units are needed because of low frequency (e.g.,

incidence of milk fever) or large variance of parameter estimates (e.g., conception rate). However, on-farm feeding trials are ill-suited to determine mechanisms, modes of action, or both. These are best addressed with more highly controlled trials performed in research herds.

On-farm feeding trials can be categorized into multi-site (**MST**) and single-site (**SST**) trials. Multi-site trials require that numerous farms be used simultaneously. They are best used to estimate the frequency of success and to identify factors that enhance or reduce this frequency. Single-site trials reduce the complexity associated with an experiment involving a multitude of locations. This reduction in complexity is done at the expense of the inference range, that is, conclusions are valid for a much narrower range of conditions. Single-site trials are best used to demonstrate success or failure at a given location, under a specific set of conditions. It is tempting to pool results of SST experiments a posteriori in an attempt to create a post facto multi-site experiment. Because different protocols are generally used across sites, and because of the partial or total confounding between sites and time, this pooling practice generally leads to erroneous conclusions regarding expected response or rate of success. Therefore, the planning of on-farm feeding trials needs to be discussed.

Planning of On-Farm Feeding Trials

As with any other research endeavor, the proper planning of on-farm feeding trials is a critical step to ensure that valid conclusions are reached. First, the experimenter must determine the objectives. These, in conjunction with availability of resources and practical constraints, determine the best experimental design. In instances in which nonconventional designs are used, the researcher must evaluate the statistical validity of the design and the a priori ability to identify parameters to be estimated. This process can be accomplished by simulating observations on a computer, making use of Monte Carlo techniques (1, 7). The simulated data are analyzed as if they were real, using the same statistical software as the one to be used once real observations will be collected. With this procedure, the researcher can identify potential problems with the planned analysis, especially those related to parameter identification and confounding of effects. Our experience indicates that the lack of planning regarding the statistical analysis of data is the single largest contributor to failure of on-farm feeding trials. At best, it is difficult, but generally impossible

to recover from a poorly designed experiment. Researchers should decide on a method of analysis with care, and, if there is any doubt about the appropriateness of a method, the aid of a statistician should be sought.

Design Consideration

There are a number of factors to consider when designing an on-farm feeding trial. The most important and most often abused are randomization, confounding of effects, identification of experimental units, replication, and validation of measuring equipment.

Randomization. Randomization must be ensured with respect to physiological status of animals, time, and location. Factors that can be considered important to the physiological status include parity, stage of lactation, pregnancy, milk production, and composition. Randomization with respect to time and location reduce the likelihood of biased estimates of treatment effects.

Confounding of effects. Many experimental designs purposely confound effects of low interest while enhancing the power of the tests for those effects of high interest (2, 6). Therefore, confounding is not necessarily wrong. But confounding is undesirable when it leads to biased tests of those effects of importance. Unfortunately, this situation often arises in poorly planned experiments in which treatment effects are inadvertently confounded with time, location, or other management factors.

An example of confounding with time occurred when one of us was asked to analyze the results of a lactating cow feeding trial. The experiment took place at a single location with two identical pens of cattle, each given a treatment. There were three levels of a feed additive fed (1×, 1.5×, and 2×). These were administered sequentially over time. The calculation of raw treatment means showed that the 2× level resulted in the lowest production. The experiment had no independent replication (pens), which prohibited a valid analysis of the data. Also, a period of heat stress started at the same time as the 2× treatment. This environmental effect, caused by time, was totally confounded with the treatment effects, making any attempt of a proper analysis completely futile.

A few years ago, one of us was asked to visit a poultry research facility that was experiencing inauspicious results during the summer season. The poultry house had an east-west orientation. A careful review of prior research protocols showed that control pens were four times more likely to be located on the

north side than on the south side. Of course, the opposite was true for the treatment pens. This resulted in a near confounding of treatments with pen locations, and the erroneous conclusion that the treatment was just not effective during the summer months. Ventilation, housing, and other factors need to be randomized across treatments.

An example of confounding by management practices happened when a feed company nutritionist investigated the effect of a feed additive on milk production. The experimenter knew that the feed additive was generally very effective at improving milk production. He wanted to demonstrate the effect to one of his clients. Only one large pen of cows was available for the demonstration. Therefore, the experimenter chose a switchback design with experimental periods of 1 mo. During the first month, cows were assigned the control diet. In the second month, they were switched to the treatment diet containing the feed additive. A month later, cows were switched back to the control. In the fourth month, cows were again assigned to the treatment diet. Milk production data were taken from monthly DHI test reports. An initial analysis showed that cows produced considerably less milk during the two periods in which they were fed the feed additive. It was later found that the farm was being tested on an a.m. and p.m. schedule in which one month all cows were sampled at the morning milking and the next month, all were sampled in the afternoon milking. The treatment effect was totally confounded with a management practice.

Identification of the experimental unit. The proper identification of the experimental unit seems to cause problems to many experimenters. Defined briefly, the experimental unit is the smallest unit to which a treatment can be applied randomly. With modern husbandry practices, it is customary to group animals in pens. Depending on how the treatment is applied, the experimental unit can be the pen or the animal.

An experiment was done using two pens of 50 calves each. Treatments were control and a feed additive fed in a TMR. The control treatment was randomly assigned to pen 1, and the feed additive treatment to pen 2. Calves were randomly assigned to pens. The investigator analyzed the data as a completely randomized design, ignoring the pen effect. A schematic analysis of variance for this experiment along with the expected mean squares (EMS) is shown in Table 1. Because there was no replication of pen within each treatment, the pen effect has no degree of freedom and is not testable. However, the EMS and pen within treatment clearly shows that the

TABLE 1. Schematic analysis of variance table with expected mean squares for an example in which 100 calves were assigned at random to two pens and treatments were applied to the pens (fixed effects of treatments; random effects of pen).

Effect	df	Expected mean squares ¹
Treatment	1	$\sigma_w^2 + k_4\sigma_{C:PT}^2 + k_5\sigma_{P:T}^2 + k_6\kappa_T$
Pen (treatment)	0	$\sigma_w^2 + k_2\sigma_{C:PT}^2 + k_3\sigma_{P:T}^2$
Calf (pen \times treatment)	98	$\sigma_w^2 + k_1\sigma_{C:PT}^2$

¹ σ^2 = Variance components, k_i = coefficients, and κ_T = the fixed effect of treatments. The appropriate denominators for F tests are pen (treatment) to test the treatment effect, and calf (pen \times treatment) for the pen (treatment) effect.

latter is the proper error term for testing the treatment effect. Because the experiment was not designed properly, the pen effect could not be estimated. The only recourse left was to assume that pens did not have an effect on results (i.e., the variance component of pen within treatment is null, $\sigma_{P:T}^2 = 0$) and use calf within pen by treatment as the error term. In doing so, the experimenter would assume 1) no pen effect, and 2) that errors within pens were independent. Both assumptions rest on a fragile terrain at best. The analysis of properly designed experiments generally shows a significant and important pen effect (14). With the improperly designed experiment, this effect is attributed to treatments. Consequently, it is likely that many experiments with pens that are not replicated have concluded erroneously to a significant treatment effect that, in fact, was solely attributable to a nontestable pen effect. Additionally, the assumption of independence of errors within pen is seldom true. Animals are generally weighed sequentially by pen. Animals within a pen are fed at the same time. The micro-environment is also more uniform within a pen than across pens. These factors and many others cause nonindependence of errors. Mixed model analyses confirm that, generally, the independence of errors within a grouping factor cannot stand (10, 11). Under these conditions, estimates of treatment effects (e.g., treated vs. control) are unbiased, but variance estimates (e.g., standard error of treated vs. control) are biased upward or downward depending on the structure of the errors.

There are instances where pens are cross-classified with treatments, resulting in a different analysis of variance table. Assume that an experiment is conducted with two pens of 50 calves each. This time, we are testing the effect of an injectable antibiotic. Half the calves in each pen are randomly assigned to a control (injection of saline) and half to a treatment (injection of antibiotic). With this design, both treatments appear in both pens. The schematic analysis of

TABLE 2. Schematic analysis of variance table with expected mean squares for an example in which 100 calves were assigned at random to two pens and treatments were applied at random to calves within pens (fixed effects of treatment; random effects of pen).

Effect	df	Expected mean squares ¹
Treatment	1	$\sigma_w^2 + k_6\sigma_{C:PT}^2 + k_7\sigma_{TP}^2 + k_8\kappa_T$
Pen	1	$\sigma_w^2 + k_4\sigma_{C:PT}^2 + k_5\sigma_P^2$
Treatment × pen	1	$\sigma_w^2 + k_2\sigma_{C:PT}^2 + k_3\sigma_{TP}^2$
Calf (treatment × pen)	96	$\sigma_w^2 + k_1\sigma_{C:PT}^2$

¹ σ^2 = Variance components, k_i = coefficients, and κ_T = the fixed effect of treatments. The appropriate denominators for F tests are treatment × pen to test the treatment effect, and calf (treatment × pen) for the treatment × pen effect.

variance table is shown in Table 2. The proper error term for testing the treatment effect is the treatment by pen interaction. This interaction can be tested with the calf within pen by treatment term. Often, the interaction effect proves to be negligible and can be pooled with the calf within pen by treatment effect, resulting in a more powerful test with an error term with 97 degrees of freedom that is used for testing the treatment effect.

At this point, it should be clear that the derivation of the expected mean squares is a critical step to insuring that valid tests are performed on a properly designed experiment. Although several textbooks present rules for figuring expected mean squares, those presented by Damon and Harvey (3) as originally stated by Henderson (9) appear to be the clearest and are as follows:

1. First assume that all classifications are random.
 - 1.1 Variance component (σ^2) values corresponding to each of the possible lines of the analysis of variance table appear in one or more of the expectations of mean squares. The subscripts on σ^2 are the same as the identification for the sources of vari-

ation. If the mean square cannot be computed because it has 0 degrees of freedom, the corresponding σ^2 is retained nevertheless.

- 1.2 A component of variance σ_w^2 appears with a coefficient of one in the expectation of every mean square.
- 1.3 In addition to σ_w^2 , the expectation of a mean square contains any σ^2 described in 1.1 that has in its subscripts all the letters denoting that mean square.
2. Determine classifications that are fixed. The expectations are the same as those calculated previously according to the rules of section 1, except that certain σ^2 are deleted according to the following rule. Any σ^2 having to the left of the colon a letter denoting a fixed classification disappears from the expectations of all mean squares not containing this letter (by definition, the colon is assumed to the right of all subscripts if there is no colon).

The coefficient (k) of a particular σ^2 in a particular mean square is calculated as N (the total number of observations) divided by the product of the n corresponding to the letters in the subscript of σ^2 . With unbalanced designs (unequal subclass frequencies) the computation of σ^2 coefficients is not straightforward and is best determined by statistical software such as SAS (13).

The rules for deriving the EMS clearly show the importance of defining whether a factor has a random or a fixed effect. In Table 3, the expectations of mean squares are shown for various models of the randomized complete block design. This table shows that the proper error term for treatments is the interaction of blocks by treatments if blocks are considered random. However, if blocks are considered as fixed, the residual (animals within blocks by treatments) is the

TABLE 3. Expectation of mean squares (EMS) with three models of randomized complete block designs.

Source	EMS of Models ¹		
	Random random	Fixed random	Fixed fixed
Treatments (T)	$\sigma_w^2 + k_4\sigma_{TB}^2 + k_5\sigma_T^2$	$\sigma_w^2 + k_3\sigma_{TB}^2 + k_4\kappa_T$	$\sigma_w^2 + k_3\kappa_T$
Blocks (B)	$\sigma_w^2 + k_2\sigma_{TB}^2 + k_3\sigma_B^2$	$\sigma_w^2 + k_2\sigma_B^2$	$\sigma_w^2 + k_2\kappa_B$
T × B	$\sigma_w^2 + k_1\sigma_{TB}^2$	$\sigma_w^2 + k_1\sigma_{TB}^2$	$\sigma_w^2 + k_1\kappa_{TB}$
Error (W)	σ_w^2	σ_w^2	σ_w^2

¹First term describes treatment in model; second term describes block in model.

² σ^2 = Variance components, k_i = coefficients, and K_j = fixed effects.

proper error term for testing treatments. Therefore, the distinction between random and fixed effects is vital to determining the proper ratio of mean squares to be used in the F tests. Traditionally, block effects have been considered fixed (e.g., 6, 2). Computation convenience to avoid mixed models was the main justification. Neter and Wasserman (12) suggested that an effect should be considered random when factor levels are not of intrinsic interest but constitute a sample from a larger population of factor levels. Damon and Harvey (3) argued that random effects are those in which the levels are considered as a random selection representing an infinite population. The same thinking was followed by Littell et al. (11). This concept that the levels must come from a *random* selection of an infinite *population* has created much confusion among researchers. At the limit, the number of pens is a finite number. Therefore, strictly speaking, pen should not be a random effect. Also, some have argued that they are certainly not selected at random. Recently, Douglass (5) suggested that the classification of random and fixed effects should be determined by the answer to the following question: would a repetition of the experiment results in estimates of the same effects? If the answer is yes, then the effect should be considered fixed; otherwise, it is random. Following this line of thinking, block effects will generally be classified as random. If commercial farms are used to conduct nonregulatory on-farm feeding trials, then the farm effects can be considered fixed or random depending on the inference range sought by the investigator. If the interest is only in those specific farms used for the trial, then farms can be considered a fixed factor. In such instances, inference is limited to the farms used in the trial. However, if the investigator's interest is on the treatment effects over the whole (large) population of farms, then the farms' effects should be considered random. In the latter case, the investigator still has to answer the question as to whether the selected farms were representative of all farms, or just a subset of those. This 'expert' judgment has an obvious and large impact on the inference range.

Replication. The concept of replication is often abused either during the design or the analysis of on-farm feeding trials. Replications are best when independent of each other and should be done generally at the level of the experimental unit. Replications within the experimental unit do not increase degrees of freedom for the F test. In an experiment in which pens are the experimental units (i.e., pens are nested within treatments), the degrees of freedom for treatments and pen within treatments (the proper error

term for treatments) are independent of the number of animals per pen. Only additional pens would alter the degrees of freedom. This does not mean, however, that the number of animals per pen has no effect on the analysis of variance. Adding animals to the pens (within managerial reason) reduces the variance of pens within treatments, with the consequence of a higher F ratio for treatment effects. This is why experiments with large pens do not require the same number of replicates to exhibit the same power as an experiment with individual animals. Federer (6) provides equations to estimate the relative efficiency of various experimental designs.

The concept of replicated experimental units and subunits should not be confused with that of replicated measurements over time or space as in the infamous repeated measures designs described by (10). The error structure of those designs is such that elaborate statistical techniques are required. Recent software such as Proc Mixed of SAS (11) should be used in those situations in which the errors are not independent and must be modeled to get accurate inference tests.

Validation of measuring equipment. Although precision should be considered in the selection of measuring equipment, accuracy (or lack of bias) should be the prime determinant. Feed scales that are mounted on trucks or wagons are particularly vulnerable because of the harsh environment in which they are generally operated on commercial farms. A protocol to verify the accuracy of feed scales using known weights with the mixer empty, half empty, and near full must be implemented at the beginning and end of trial at the very minimum. The calibration of milk metering and sampling devices as well as those of any other instrument (pH meters, moisture testers, etc.) must be ensured on a regular basis. No statistical technique can help the investigator recover from unknown, biased measurements.

Along with validation of measuring equipment, a properly designed on-farm feeding trial should insure that feed samples are properly collected, analyzed to establish that the feeding programs were as designed, and preserved in case there is a future need for determining whether treatments were administered as expected. Management logs must be designed to allow for easy but accurate maintenance and inspection.

CONCLUSIONS

When properly designed, on-farm feeding trials can serve a useful purpose in the scientific process.

However, they are often plagued with problems that largely result from poor planning. Investigators often fail to account for environmental changes over time, improperly assign animals to treatment, fail to replicate treatments across experimental units, neglect on-farm oversight, fail to analyze feedstuffs to confirm that the feeding program was as designated, and ignore the calibration of measuring equipment. Such trials waste resources, provide faulty information, and, worse, lead to erroneous management decisions.

REFERENCES

- 1 Carson, E. F., C. Cobelli, and L. Finkelstein. 1983. *The Mathematical Modeling of Metabolic and Endocrine Systems*. John Wiley & Sons, New York, NY.
- 2 Cochran, W. G., and G. M. Cox. 1957. *Experimental Designs*. 2nd ed. John Wiley & Sons, New York, NY.
- 3 Damon, R. A., and W. R. Harvey. 1987. *Experimental Design, ANOVA, and Regression*. Harper & Row Publ., New York, NY.
- 4 DeGroot, M. H. 1970. *Optimal Statistical Decisions*. McGraw-Hill, New York, NY.
- 5 Douglass, L. W. 1998. Using mixed models: an easier and safer linear models procedure. *J. Dairy Sci.* 81(Suppl. 1):314. (Abstr.)
- 6 Federer, W. T. 1955. *Experimental Design*. The MacMillan Co., New York, NY.
- 7 Fishman, G. S. 1978. *Principles of Discrete Event Simulation*. John Wiley & Sons, New York, NY.
- 8 Galligan, D. T., W. Chalupa, and C. F. Ramberg. 1991. Application of type I and type II errors in dairy farms management decision making. *J. Dairy Sci.* 74:902-910.
- 9 Henderson, C. R. 1969. Pages 1-55 *in* Design and Analysis of Animal Husbandry Experiments. Am. Soc. Anim. Prod.
- 10 Littell, R. C., P. R. Henry, and C. B. Ammerman. 1998. Statistical analysis of repeated measures data using SAS procedures. *J. Anim. Sci.* 76:1216-1231.
- 11 Littell, R. C., G. A. Millikan, W. W. Straup, and R. D. Wolfinger. 1996. *SAS System for Mixed Models*. SAS Inst., Inc., Cary, NC.
- 12 Neter, J., and W. Wasserman. 1974. *Applied Linear Statistical Models*. Richard D. Irwin, Inc., Homewood, IL.
- 13 SAS/STAT® User's Guide, Version 6 Edition. 1990. SAS Inst., Inc., Cary, NC.
- 14 Snedecor, G. W., and W. G. Cochran. 1980. *Statistical Methods*. 7th ed. The Iowa State Univ. Press, Ames.