J. Dairy Sci. 99:1–9
http://dx.doi.org/10.3168/jds.2016-11516

Short communication: On recognizing the proper experimental unit in animal studies in the dairy sciences

Nora M. Bello,*[1] Matthew Kramer,† Robert J. Tempelman,‡ Walter W. Stroup,§ Normand R. St-Pierre,# Bruce A. Craig,‖ Linda J. Young,¶ and Edward E. Gbur**
*Department of Statistics, Kansas State University, Manhattan 66506
†USDA, Agricultural Research Service, Beltsville, MD 20705
‡Department of Animal Science, Michigan State University, East Lansing 48824
§Department of Statistics, University of Nebraska, Lincoln 68583
#Department of Animal Sciences, The Ohio State University, Columbus 43210
‖Department of Statistics, Purdue University, West Lafayette, IN 47907
¶National Agricultural Statistics Service, Washington, DC 20250
**Agricultural Statistics Laboratory, University of Arkansas, Fayetteville 72701

ABSTRACT

Sound design of experiments combined with proper implementation of appropriate statistical methods for data analysis are critical for producing meaningful scientific results that are both replicable and reproducible. This communication addresses specific aspects of design and analysis of experiments relevant to the dairy sciences and, in so doing, responds to recent concerns raised in a letter to the editor of the *Journal of Dairy Science* regarding journal policy for research publications on pen-based animal studies. We further elaborate on points raised, rectify interpretation of important concepts, and show how aspects of statistical inference and elicitation of research conclusions are affected.
Key words: experimental unit, replication, observational unit, hierarchical data structure, pen

Short Communication

Sound design of experiments and proper implementation of appropriate statistical methods for data analysis are critical for producing meaningful scientific results that are both replicable and reproducible (Milliken and Johnson, 2009). First, consider the concept of a "statistical unit," as proposed by Robinson (2016) in a recent Letter to the Editor in the *Journal of Dairy Science*, a term that is decidedly vague and lacks a universal definition in the mainstream design of experiments literature, particularly for agricultural applications (Kuehl, 2000; Littell et al., 2006; Casella, 2008; Milliken and Johnson, 2009; Stroup, 2013). Instead, let us define the "experimental unit" and the "observational unit," both formally and in the specific context of the dairy sciences. The leading literature in design of experiments defines the experimental unit, also called the unit of replication, as *the smallest entity that is assigned independently of all other units to a particular treatment*; the word *independent* is key to this definition (Kuehl, 2000; Littell et al., 2006; Casella, 2008; Milliken and Johnson, 2009; Stroup, 2013). Experimental units are often assumed to be "exchangeable," a statistical term that implies that the units do not differ in any fundamental way, so that reliable inferences would be obtained regardless of which treatment was assigned to each unit.

In the dairy sciences, individual cows can sometimes serve as experimental units; for example, if treatments were different types or doses of antibiotics individually injected to treat mastitis. Even then, cows may still be housed together in pens but individual cows within a pen are randomly assigned to different treatments. In dairy nutrition, it is often of interest to compare diets that, for logistical reasons, are commonly fed (i.e., randomly assigned) to pens, such that all cows in the same pen are offered the same diet. For example, if one wanted to compare 2 diets, one could design an experiment by randomly assigning diets A and B each to a different random set of pens, with each pen holding several cows. In this case, the pen is clearly the experimental unit. If 2 pens receiving different diets showed any difference in outcome, we would not know whether this difference was due to the intended diet effect, a confounded pen effect, or a combination of both effects. To effectively separate diet effects from pen effects would require more pens; that is, diets need to be replicated to multiple pens. How many more pens? This is a question of statistical power and depends on how large the diet effect is expected to be, how variable observations from pens fed the same diet are, and how

Received May 26, 2016.
Accepted July 31, 2016.
[1] Corresponding author: nbello@ksu.edu

this variability partitions into pen-level (i.e., between-pen) variability and cow-level (i.e., within-pen) variability. For further details on statistical power in the context of the dairy sciences, the reader may refer to Tempelman (2009).

Distinct from an experimental unit, to which a treatment is independently applied, is the concept of an *observational unit*, also known as the sampling unit. This distinction is recognized in the response to Robinson (2016) by Lamberson (2016). An observational unit is defined as *the physical entity on which an outcome of interest is measured in an experiment* (Kuehl, 2000; Casella, 2008). In many simple designs, experimental units and observational units are synonymous; that is, they can be matched to the same physical entity (Kuehl, 2000; Littell et al., 2006; Stroup, 2013). This was true in the prior example when assessing the effect of antibiotic treatments individually injected and can also be true for the diet example if the outcome of interest were measured at the pen level (e.g., total intake for the pen or total time spent feeding for all animals in a pen). If pen is the entity that is both independently assigned to treatment and measured for outcome, then pen serves as both the experimental unit and the observational unit. On the other hand, if the outcome of interest in the diet example was measured on individual cows in each pen, say milk yield, one encounters a natural "gap" or "mismatch" between the entity independently assigned to treatment (i.e., pen) and the entity measured (i.e., individual cow within a pen). This is an example of a nested design structure: the pen is nested within a treatment and the individual cow is nested within a pen, thereby creating a hierarchical structure in the data.

A hierarchical data structure refers to a configuration of the data where observations are not mutually independent but rather have a correlation structure imposed by the experimental design. In our dairy example with diets applied to pens, pens consist of individual cows but these animals are not mutually independent and, consequently, neither are their observations. Specific biological reasons to explain lack of independence of observations collected on cows within a pen are context specific. In the dairy sciences, one can often anticipate within-pen dynamics; for instance, differential feed access due to social behavior (e.g., dominance) or management practices (i.e., feed mixing). Notably, this kind of correlation between observations from cows within a pen is different from a general "pen" effect, which may be due, for instance, to pen size, condition of the substrate, or shade availability, to name a few. It is precisely due to this correlation (i.e., lack of independence) between cow-level observations that it is not possible to separate diet effects from pen effects in a nutrition study conducted on only 2 pens, regardless of the number of cows in each pen. Whenever observational units are nested within an experimental unit, as is the case here, the observational units are commonly referred to as subsamples, pseudoreplicates, or technical replicates (Casella, 2008) to indicate that these observations are correlated and thus do not constitute true independent replication. Data structures such as these are common in the animal sciences; examples include multi-farm studies, groups of animals entering a study in weekly clusters, or repeated observations collected over time on individual animals (i.e., test-day milk yield). Hierarchical data structures, and thus underlying correlations between observations, can often be recognized as nesting or blocking in the experimental design of a study. Both nesting and blocking are common elements of design in dairy trials; thus, it is not surprising that experimental units are often separate physical entities from observational units in dairy science experiments.

We emphasize: *experimental units are defined in terms of independent treatment assignments whereas observational units are defined in terms of outcome measurements.* These are clearly different definition criteria. As such, observations do not necessarily represent replications. However, observational units are usually contained within experimental units (Stroup, 2013), which in turn determine the amount of replication of a given experiment. As a side note, a potential exception is a repeated-measures design, and this depends on whether one labels the observational unit to be an individual cow or an individual cow at a specific time point—here, labels are less important than the concept that repeated measures on the same cow are mutually correlated. Even so, replication implies an independent repetition of a basic experimental component, such as a treatment, and is considered a prime requisite for valid and reliable experimental inference (Kuehl, 2000; Casella, 2008). The rationale to support true replication as a requisite for valid experiments is well explained by Kuehl (2000), including the following: (1) results are reproducible, at least under the specified experimental conditions; (2) results are not aberrant realizations of an experiment due to unforeseen circumstances; and (3) variability between experimental units defining experimental error is properly estimated and thus subsequent hypothesis tests are properly calibrated.

To be able to identify hierarchical data structure; that is, when independent replication occurs and when it does not, it is most important to understand the complete process involved in collecting data and carrying out a study. This understanding is also critical to adequately specify the statistical model for data analysis. For illustration purposes, consider alternative layouts for a general 3 × 3 Latin square design consisting of 3