Animal Feed Science and Technology 171 (2012) 1–5

Contents lists available at SciVerse ScienceDirect

Animal Feed Science and Technology

journal homepage: www.elsevier.com/locate/anifeedsci

Editorial

Use of replicates in statistical analyses in papers submitted for publication in Animal Feed Science and Technology

ARTICLE INFO

Keywords:
Replicates
Statistical analyses

ABSTRACT

This editorial is an annex, essentially an amplification of earlier Animal Feed Science and Technology (AFST) editorials which focused on analytical and statistical issues of papers submitted for publication in AFST. This amplification is needed because, since publication of those editorials, there has been a sharp increase in papers submitted to AFST (particularly *in vitro* gas, ruminal *in sacco*, continuous culture fermenter and mini-silo studies), in which there have been substantive disagreements among authors, reviewers and editors as to what, exactly, constitutes acceptable statistical replicates as opposed to pseudo replicates. In this editorial, the Co-Editors in Chief of AFST provide clarification of their view relative to use of analytical observations in statistical analyses in papers submitted for consideration for publication in AFST. If the objective is to compare feeds and from the results, make inferences to populations, only multiple samples of each feed represent an acceptable feed base. This suggests that means comparisons based on repeated assays of the same sample, whether by chemical, physical or microbiological methods, will almost certainly be rejected by the CEIC.

1. Introduction

This editorial is an annex, essentially an amplification of earlier editorials (Udén et al., 2005; Robinson et al., 2006) which focused on analytical and statistical issues, respectively, in manuscripts submitted for consideration for publication in Animal Feed Science and Technology (AFST). Since the publication of those editorials, there has been a sharp increase in the number of papers submitted to AFST, particularly *in vitro* gas, ruminal *in sacco*, continuous culture fermenter and mini-silo studies, in which there have been substantive disagreements among authors, reviewers and editors as to what, exactly, constitute acceptable statistical replicates. In many cases, pseudo replicates, in the view of an AFST Co-Editor in Chief (CEIC), were used by authors in the statistical analysis. As the definition of replicates is not always apparent in the 'Materials and methods' section of the original manuscript, the AFST CEIC have often had to ask, occasionally repeatedly, for more information regarding experimental and statistical design details before the statistical procedures followed by the authors became fully clear.

In this editorial, the CEIC of AFST provide clarification of their view as to use of replicated analyses in statistical analyses and how to describe the data in manuscripts submitted for consideration for publication in AFST. Note that 'replicate' in this editorial is not used synonymously with 'statistical unit'. Only when preceded by 'true' do they mean the same thing.

2. Experimental objectives of a study

The objective of a study is crucial to its statistical analysis (Robinson et al., 2007), as it determines how and what type of statistical analysis should be completed. In many cases, manuscripts submitted to AFST have objectives which are not

Abbreviations: AFST, Animal Feed Science and Technology; CEIC, Co-Editors in Chief; CV, coefficient of variation; MPE, mean prediction error; MRSS, mean residual sums of squares; MSPE, mean square prediction error.

0377-8401/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.anifeedsci.2011.10.008

sufficiently explicit. Often, authors claim universality of objectives, whereas the experimental conditions and the sample sets are unique and inferences are limited. Objectives must always be clear and specific in order to inform the reader as to the purpose of the study. An example of a specific objective might be:

> "To compare in vitro digestibility of first-cut timothy (cv Grindstad) and white clover (cv Ramona) harvested at four maturities grown in northeast Wisconsin on a clay soil during 2009".

In this example, the reader is not misled into making inferences to grasses and clovers in general. However, acceptance of such an objective as a valid basis for a study is not guaranteed and will be based on its novelty, scientific value and applicability outside of the confines of the experiment. In this case, multiple locations, cuts, varieties and/or years may be required for manuscript acceptance in order to provide appropriate representation of the plants under study.

Experimental design, statistical analyses and experimental objectives must all be coherent. In some cases, all of these are correct in a formal sense but, if the objectives represent non-issues (which is often a matter of scientific opinion), then the study is not publishable. Comparisons of plant species based on single or pooled samples, treatments applied to single batches of feeds, or validations of one unreliable method against another unreliable method are examples of objectives which are likely to be rejected.

In the context of this editorial, we examine three main types of investigations with the objectives: (i) investigation of a method, (ii) comparison of two or several methods, and (iii) comparison of treatments, feed sources, plants and other similar materials.

3. Studies to evaluate a single method

The objective of a single method evaluation study is to investigate methodological errors within an assay without information on true composition. Analyses may be with or without the possibility for calibration. While precision is the primary objective, accuracy may also be investigated. Analyses are generally replicated within and over time and all observations are used in statistical analysis. Total variance measured in a detailed study can be portioned into what is defined as repeatability or intra-assay errors (*i.e.*, an assay repeated under the same operating conditions and time), and reproducibility errors. Reproducibility errors are generally of higher interest than repeatability errors, but they have several meanings as they could be inter-laboratory errors, errors dependant on time (run), errors among technicians or other similar types of errors. However, if the method is to apply to a wide range of feeds, it is critical that they are well represented in the study. A good example is the neutral detergent fibre method (Mertens, 2002) where inter-laboratory errors were found to be higher for starchy feeds than for forages due to incomplete removal of starch and ensuing filtration problems for starchy feeds.

Dealing with analytical errors within a laboratory requires use of standards of different kinds, and is outside the scope of this editorial.

4. Studies to compare methods

In this type of study, the objective is to compare a test method (*i.e.*, new or revised) to a reference method or, occasionally, a single method against "known" values. It is critical in methods comparison studies that there be a reference method and/or a biologically sensible way to evaluate the results. However, in situations where no accurate method or "gold" standard exists, there may be scope for methods comparisons based on sensitivity to treatment changes and/or ranking of feeds relative to biological responses.

As mentioned in Section 3, it is important that feeds are well represented in such studies. Repeated analyses (on the same sample) should not be a part of the design of this type of study in its pure form. Partitioning of errors into mean bias, slope and disturbance in these types of studies can be useful. One such technique has been outlined by Theil (1966) and Bibby and Toutenberg (1977), as well as more recently by Dhanoa et al. (1999). Calibration errors should only include disturbance but, in method validation, bias and slope errors also occur in various proportions dependant on the origin of the test set of feed samples. Bias and slope errors should prompt further investigation into why sample sets differ. Errors due to disturbance should be evaluated relative to the practical applications of the method, which may be scientific, advisory, safety or other. Note that mean square prediction error (MSPE) cannot be directly compared with mean residual sums of squares (MRSS) from regression, as the former is calculated as the sum of the squared differences divided by 'n', while the latter is calculated from the squared deviations from regression divided by $n - 2$. Thus, they only become similar at high values of 'n', and as the slope approaches 1. A key issue is how to present the reliability of the method. The coefficient of variation (CV) and mean prediction error (MPE) are calculated from MRSS and MSPE, respectively, and can be used as a measure of proportional error. However, if a limited data set with only low values is used, CV and MPE will be high as analytical errors are considered not to be proportional to recorded values but equal as implicit in any standard analysis of variance. Thus, if low values are not of particular interest, it may be advisable to report the absolute error and relate that error to the expectations of the method relative to its likely practical application.

5. Studies to compare treatments

The objectives of such experiments may be to examine differences among feeds or effects of physical, chemical or microbiological treatments of feeds. Problems during review tend to occur when treatment and measurement variability is confused by authors.

The reason is that individual feeds or diets contain no variability in terms of source; only the evaluation method itself will cause variability. This variability will be higher if animals are used and lower if chemical analyses are employed.

5.1. Laboratory and in sacco methods

Chemical or *in vitro* studies are relatively easy to replicate in terms of feed sources, and the experimentalist should resist the temptation to save money by pooling samples from individual plots or trees, which is too often the case, thereby eliminating sample variation in the study. Plant variety testing utilizing chemical and/or *in vitro* analyses requires observations from multiple fields/plots/trees at several locations and, potentially, collected over several years. Animal nutritionists are often satisfied with single samples from one location harvested once in a single year as the basis of studies. This is a serious limitation for plant evaluations and will seriously impact the extent of the conclusions which can be drawn from reported differences among feeds. With this in mind, authors should avoid limited studies of this nature as AFST acceptance of unreplicated plant species comparisons is highly unlikely when the only replication consists of repeated assays (either chemical or *in vitro*) of the same material. This means that fields/plots/trees, which will themselves be a source of variation, must be replicated in order to obtain a sufficient number of statistical units (as discussed above and in Robinson et al., 2006). Treatments such as feed processing, field applications of chemical, microbiological, physical and similar nature, or in the form of inocula from animals fed contrasting diets must also be replicated similar to that described above in order to achieve true replication. Sometimes studies are conducted where known amounts of defined (an AFST requirement) and consistent chemicals or biological additives are applied to feeds or forages under laboratory conditions. In such cases, it is generally of less interest to repeat the applications for statistical purposes. However, when additive level effects and feed by additive interactions are likely to occur, multiple feeds and/or levels must be used as there is no other way to obtain a sufficient number of statistical units.

Repeated chemical and *in vitro* analyses on the same sample, or data from replicated mini-silos are generally necessary. However, to be consistent with objectives mentioned above, values should be averaged before being analyzed statistically. The number of analytical replicates needed will depend on variability of the technique employed and is a matter of judgment by the experimentalist. In the case of *in sacco* and *in vitro* measurements, there may specific reasons to repeat the analyses on different days (*i.e.*, multiple runs). This may be due to lack of space or to safeguard against errors which are particularly associated with runs (*e.g.*, inocula quality problems). However, to include 'run' in the statistical model, either as a fixed or random factor, to evaluate treatment effects is incorrect because the CEIC regard runs as repeated observations of 'identical' treatments. In well managed laboratories, using standard protocols, run differences are normally corrected by including appropriate standards or, if judged necessary, by repeating the run. To reduce the effects of unusual rumen inocula or rumen environments, the CEIC suggest that, normally, at least three animal sources be used for *in sacco* and *in vitro* measurements.

A completely different situation occurs when feeds or diets have quantitative traits, such as its protein quality or tannin content. In such cases, the objective should not be to compare feeds, but to compare traits similar to dose level responses and analyzed as outlined by Robinson et al. (2006).

5.2. In vivo studies

It should be self evident that very few large animal studies could be published if multiple sources of feeds/diets for each treatment were required for manuscript acceptance in order to meet objectives such as those mentioned earlier in Section 5. However, bales of hay or silage, or concentrate feeds from multiple sources offer possibilities to introduce true treatment replications into *in vivo* studies and the CEIC encourage research efforts using this approach. With animals requiring small amounts of feed (*e.g.*, chickens), this is not only possible but often required to capture variability within feeds. In large animal studies when this is practically impossible, the CEIC of AFST will insist on caution in interpretation of such results, particularly when forages are investigated because of their high compositional variability.

For feeds having quantitative traits, the same logics apply as outlined in Section 5.1. These studies are often difficult to design as it is critical that as many as possible of all other diet components are the same in order to prevent confounding of feed level with chemical composition. Generally, only differences in minor components or effects attributable to single feed components (*e.g.*, phosphorous availability affected by phytin level) are reasonably unequivocal. This is a frequent issue in feed substitution studies, and often a reason for rejection of papers since it becomes unclear if the substitutions in feeds are being made to examine a change in nutrient (in cases where efforts are made to equalize all other nutrients, as noted above) or whether it is a simple comparison of substitution of feed A for feed B.

Meta analysis may provide a solution to inclusion feed variability by use of multiple studies with similar hypotheses and should be particularly appropriate when numerous contradictory studies are found in the literature. There are a number of pitfalls in meta analysis which have been outlined by, for example, Phillips (2005). The use of mixed model procedures is favored by St-Pierre (2001) with 'Study' as a random effect for giving more biologically meaningful interpretations. In

SAS® Tools for Meta-Analysis (SAS, 2011), Hamer and Simpson give a simple outline of meta analysis with more details presented on analysis and data acquisition are presented in the papers of St-Pierre (2001), Sauvant et al. (2008) and Lean et al. (2009). As mixed models are most likely to be used in meta analyses, it should be noted that standard error of difference should be reported as the standard error of the mean cannot be easily translated into probability for treatment differences when random factors are in the model (Littell et al., 2006). Some general linear models procedures also offer the choice of random factors but the *F* test provided by the software may be incorrect (Littell et al., 2006). Authors are therefore required to examine the F tests provided by the particular software chosen.

6. Description of data used for statistical analysis

Inclusion of replicates in a statistical model will depend on the objectives, as discussed in Section 2. However, to allow reviewers and the CEIC to fully evaluate the statistical analysis, there is an absolute requirement for an unambiguous and complete description of how the total available observations were used in the statistical model. An example of this is:

> Five starch levels (L) and pooled rumen sources of inocula (INO) from two groups of three cows fed either a high or low protein level were examined in a 36 h *in vitro* batch culture. Five different forages and a single batch of purified maize starch were available. These treatment combinations were replicated twice per batch and repeated four times (runs) over a period of 2 weeks. The total number of observations was 5 (L) × 2(INO) × 5(forages) × 2(reps.) × 4(runs) = 400 but, after averaging batch and run replicates, the remaining 50 observations were subjected to analysis of variance using the model: . . ..

7. Conclusions

Animal Feed Science and Technology is a scientific journal with a wide international readership as well as a wide variety of types of experiments that are published within it. The statistical expertise (and views) of authors, reviewers, CEIC and readers varies widely.

One important criterion for acceptance of a manuscript submitted to AFST is whether the objectives were met by the design of the study. Evaluation of feeds by any method, including use of animals, must be based on the concept of what they represent as they are drawn from a population, and that population must be defined. A single sample drawn from a population cannot contribute sufficient information to warrant publication. In addition, a decision on rejection or acceptance of a manuscript can only be based on a clear description of the sample base and how the observations have been used in statistical analysis. It is not an infrequent occurrence that manuscripts submitted for consideration for publication in AFST fail to clear this bar which leads to rejection, or an excessive number of revisions which may ultimately lead to rejection, after a great deal of reviewer, CEIC and author time has been expended.

It is not possible that a single editorial can deal with all possible statistical replicate issues in all types of experiments in all types of manuscripts submitted for consideration for publication in AFST. Thus, the purpose of this editorial was to outline principles of statistical replication, as well as highlight some of the more common issues that cause authors, reviewers and CEIC to disconnect during the review process. If authors wish to obtain an AFST evaluation of the potential acceptability of statistical replicates prior to initiating an experiment, they are encouraged to contact any of the CEIC for an opinion either before initialization of the study or prior to its submission.

Acknowledgements

The editors are greatly indebted to James G. Fadel and Thomas R. Famula, UC Davis, California, USA and to Jim Craigon, University of Nottingham, UK for constructive criticism of the manuscript.

References

Bibby, J., Toutenberg, H., 1977. Prediction and Improved Estimation in Linear Models. John Wiley & Sons, New York, NY, USA (Chapter 1).
Dhanoa, M.S., Lister, S.J., France, J., Barnes, R.J., 1999. Use of mean square prediction error analysis and reproducibility measures to study near infrared calibration equation performance. J. Near Infrared Spectrosc. 7, 133–143.
Lean, I.J., Rabiee, A.R., Duffield, T.F., Dohoo, I.R., 2009. Use of meta-analysis in animal health and reproduction: methods and applications. J. Dairy Sci. 92, 3545–3565.
Littell, R., Milliken, G., Stroup, W., Wolfinger, R., Schabenberger, O., 2006. SAS for Mixed Models, 2nd edition. SAS Press.
Mertens, D.R., 2002. Gravimetric determination of amylase-treated neutral detergent fibre in feeds with refluxing beakers or crucibles: collaborative study. J. AOAC Int. 85, 1217–1240.
Phillips, C.J.C., 2005. Meta-analysis—a systematic and quantitative review of animal experiments to maximise the information derived. Anim. Welf. 14, 333–338.
Robinson, P.H., Wiseman, J., Udén, P., Mateos, G.G., 2006. Some experimental design and statistical criteria for analysis of studies in manuscripts submitted for consideration for publication. Anim. Feed Sci. Technol. 129, 1–11.
Robinson, P.H., Udén, P., Wiseman, J., Mateos, G.G., 2007. Some suggestions and guidelines for preparation of manuscripts for submission for consideration for publication. Anim. Feed Sci. Technol. 134, 181–188.
SAS, 2011. SAS® Tools for Meta-Analysis., http://www2.sas.com/proceedings/sugi27/p 250-27.pdf.
Sauvant, D., Schmidely, P., Daudin, J.J., St-Pierre, N.R., 2008. Meta-analyses of experimental data in animal nutrition. Animal 2, 1203–1214.
St-Pierre, N.R., 2001. Invited review: integrating quantitative findings from multiple studies using mixed model methodology. J. Dairy Sci. 84, 741–755.

Theil, H., 1966. Applied economic forecasting. In: Theil, H. (Ed.), Studies in Mathematical and Managerial Economics, vol. 4. North Holland Publishing Co, Amsterdam, The Netherlands.
Udén, P., Robinson, P.H., Wiseman, J., 2005. Use of detergent system terminology and criteria for submission of manuscripts on new, or revised, analytical methods as well as descriptive information on feed analysis and/or variability. Anim. Feed Sci. Technol. 118, 181–186.

P. Udén*
Department of Animal Nutrition and Management, Swedish University of Agricultural Sciences, Kungsängen Research Centre, 753 23 Uppsala, Sweden

P.H. Robinson
Department of Animal Science, University of California, Davis, CA 95616-8521, USA

G.G. Mateos
Departamento de Produccion Animal, Universidad Politécnica de Madrid, Cuidad Universitaria, 28040 Madrid, Spain

R. Blank
Institute of Animal Nutrition and Physiology, Christian-Albrechts-University, 24098 Kiel, Germany

* Corresponding author. Tel.: +46 018672058; fax: +46 18672946.
E-mail address: peter.uden@slu.se (P. Udén)

24 October 2011

24 October 2011