



Meta-analyses of experimental data in the animal sciences

Normand Roger St-Pierre

1 - Department of Animal Sciences, The Ohio State University, 2029 Fyffe Rd., Columbus, OH-43210, USA. st-pierre.8@osu.edu

ABSTRACT - In certain areas of animal research, such as nutrition, quantitative summarizations of literature data are periodically needed. In such instances, statistical methods dealing with the analysis of summary data (generally from the literature) must be used. These methods are known as meta-analyses. The implementation of a meta-analysis is done in several phases. The first phase defines the study objectives and identifies the criteria for selecting prior publications to be used in the construction of the database. Publications must be scrupulously evaluated for their quality before being entered into the database. During this phase, it is important to carefully encode each record with pertinent descriptive attributes (experiments, treatments, etc.) to serve as important reference points later on. Statistically, databases from literature data are inherently unbalanced, leading to considerable analytical and interpretation difficulties. Missing data are frequent, and data are not the outcomes of a classical experimental system. A graphical examination of the data is useful in getting a global view of the system as well as to hypothesize specific relationships to be investigated. This phase is followed by a statistical study of the meta-system using the database previously assembled. The statistical model used must follow the data structure. Variance decomposition must account for inter- and intra-study sources; dependent and independent variables must be identified either as discrete (qualitative) or continuous (quantitative). Effects must be defined as either fixed or random. Often observations must be weighed to account for differences in the precision of the reported means. Once model parameters are estimated, extensive analyses of residual variations must be performed. The roles of the different treatments and studies in the results obtained must be identified. Often, this requires returning to an earlier step in the process. Thus, meta-analyses have inherent heuristic qualities that can guide in the design of future experiments as well as aggregating prior knowledge into a quantitative prediction system.

Key Words: meta-analysis, mixed models, nutrition

Abbreviation key: DMI = Dry matter intake, GLM = Generalized linear model, GLMM = Generalized linear mixed model, NDF = Neutral detergent fiber, SAS = Statistical Analysis System, SEM = Standard error of the mean, VFA = Volatile fatty acids, VIF = Variance inflation factors.

Introduction

The research environment in the animal sciences, especially nutrition, has markedly changed in recent years. In particular, there has been a noticeable increase in the number of publications, each containing an increasing number of quantitative measurements. Meanwhile, treatments often have smaller effects on the systems being studied than in the past. Additionally, controlled and non-controlled factors, such as the basal plane of nutrition, vary from study to study, thus requiring at some point a quantitative summarization, an integration of prior research.

Fundamental research in the basic animal science disciplines generates results that generally

are at a much lower level of aggregation than those of applied research (organs, whole animals), thus supporting the necessity of integrative research.

Research stakeholders, those who ultimately use the research outcome, increasingly want more quantitative knowledge, particularly on animal response to diet. Forecasting and decision support softwares require quantitative information as inputs. Lastly, research prioritization by public funding agencies may force abandoning active research activities in certain areas. In such instances, meta-analyses can still support discovery-type activities from aggregating results from published literature.

The objectives of this paper are to describe the application of meta-analytic methods to animal

nutrition studies, including the development and validation of literature derived databases, and the quantitative statistical techniques used to extract the quantitative information.

Definitions and nature of problems

Limits to Classical Approaches

Results from a single classical experiment cannot be the basis for a large inference space because the conditions under which observations are made in a single experiment are forcibly very narrow, *i.e.*, specific to the study in question. Such studies are ideal to demonstrate cause and effect, and to test specific hypothesis regarding mechanisms and modes of action. In essence, a single experimentation measures the effects of one or a very few factors while maintaining all other factors as constant as possible. Often, experiments are repeated by others to verify the generality and repeatability of the observations that were made, as well as to challenge the range of applicability of the observed results and conclusions. Hence, it is not uncommon that over time, many studies are published even on a relatively narrow subject. In this context, there is a need to summarize the findings across all the published studies. Meta-analytic methods are concerned with how best to achieve this integration process.

The classical approach to synthesizing scientific knowledge has been based on qualitative literature reviews. A limitation of this approach is the obvious subjectivity involved in the process. The authors subjectively weigh outcomes from different studies. Criteria for the inclusion or non-inclusion of studies are poorly defined. Different authors can draw dramatically different conclusions from the same initial set of published studies. Additionally, the limitation of the human brain to differentiate the effects of many factors grows with the number of publications involved.

Definitions and objectives of meta-analyses

Meta-analyses use objective, scientific methods based on statistics to summarize and quantify knowledge acquired through prior published research. Meta-analytic methods were initially developed in psychology, medicine and social sciences a few decades ago. In general, meta-analyses are conducted for one of the following four objectives:

- For ***Global hypothesis testing***, such as testing for the effect of a certain drug or of a feed additive using the outcomes of many publications that had as an objective the testing of such effect. This was by far the predominant objective of the first meta-analyses published (Mantel & Haenszel, 1959; Glass, 1976). Early on, it was realized that many studies lacked statistical power for statistical testing, so that the aggregation of results from many studies would lead to much greater power (hence lower type II error), more precise point estimation of the magnitude of effects, and narrower confidence intervals of the estimated effects.
- For ***Empirical modeling of biological responses***, such as the response of animals to nutritional practices. Because the data extracted from many publications cover a much wider set of experimental conditions than those of each individual study, conclusions and models derived from the whole set have a much greater likelihood of yielding relevant predictions to assist decision-makers. There are numerous examples of such application of meta-analytical methods in recent nutrition publications, such as the quantification of the physiological response of ruminants to types of dietary starch (Offner & Sauvant, 2004), grain processing (Firkins *et al.*, 2001) and rumen defaunation (Eugene *et al.*, 2004). Others have used meta-analyses to quantify in situ starch degradation (Offner *et al.*, 2003), and microbial N flow in ruminants (Oldick *et al.*, 1999; St-Pierre, 2003).
- For ***collective summarizations of measurements that only had a secondary or minor role in prior experiments***. Generally, results are reported with the objective of supporting the hypothesis related to the effect of one or a few experimental factors. For example, ruminal VFA concentrations are reported in studies investigating the effects of dietary starch, or forage types. None of these studies have as an objective the prediction of ruminal VFAs. But the aggregation of measurements from many studies can lead to a better understanding of factors controlling VFA

concentrations, or allow the establishment of new research hypotheses. A meta-analysis of ruminal liquid flow rates allowed the identification of an indirect criterion to saliva production and buffer recycling, which criterion is linked to ruminal conditions (Sauvant & Mertens, 2000).

- In mechanistic modeling, for *parameter estimates and estimates of initial conditions of state variables*. Mechanistic models require parameterization, and meta-analyses offer a mechanism of estimation that makes parameter estimation more precise and more applicable to a broader range of conditions. Meta-analyses can also be used for external model validation (Sauvant & Martin, 2004), or for a critical comparison of alternate mechanistic models (Offner & Sauvant, 2004).

Types of Data and Factors in Meta-Analyses

As in conventional statistical analyses, dependent variables in meta-analyses can be of various types such as binary [0, 1] (*e.g.*, for pregnancy), counts or percentages, categorical-ordinal (good, very good, excellent), and continuous, which is the most frequent type in meta-analyses related to nutrition.

Independent factors (or variables) have either a fixed or random effects on the dependent variable of interest (McCulloch & Searle, 2001). In general, factors related to nutrition (grain types, DMI, etc) should be considered as fixed effects factors. The *study* effect can either be considered as random or fixed. If a dataset comprised many individual studies from multiple research centers, the *study* effect should be considered random because each study is conceptually a random outcome from a large population of studies to which inference is to be made (St-Pierre, 2001). This is especially important if the meta-analysis has for objective the empirical modeling of biological responses, or the collective summarizations of measurements that only had a secondary or minor role in prior experiments, because it is likely that the researcher in those instances has a targeted range of inference much larger than the limited conditions represented by the specific studies. There are instances, however, where each experiment can be considered as an outcome each from a different

population. In such instances, the levels of *study* or *trial* are in essence arbitrarily chosen by the research community, and the *study* effect should then be considered fixed. In such instance, the range of inference for the meta-analysis is limited to the domain of the specific experiments in the dataset. This is of little concern if the objective of the meta-analysis is that of global hypothesis testing, but it does severely limit the applicability of its results for other objectives.

Difficulties Inherent to the Data

The meta-analytic database is best conceptualized with rows representing treatments, groups or lots, while the columns consist of the measured variables (those for which least-squares means are reported) and characteristics (class levels) of the treatments or trials. A primary characteristic of most meta-analytic databases is the large frequency of missing data in the table. This reduces the possibility of using large multi dimensional descriptive models, and generally forces the adoption of models with a small subset of independent variables. Additionally, the underlying data design, referred as the meta-design, is not determined prior to the data collection as in classic randomized experiments. Consequently, meta-analytic data are generally severely unbalanced and factor effects are far from being orthogonal (independent). This leads to unique statistical estimation problems similar to those observed in observational studies, such as leverage points, near collinearity, and even complete factor disconnectedness, thus prohibiting the testing of the effects that are completely confounded with others.

An example of factor disconnectedness is shown in Table 1, where two factors each taking three possible levels are investigated. In this example, factor A and B are disconnected because one cannot join all bordering pairs of cells with both horizontal and vertical links. Consequently, the effect of the third level of A cannot be estimated separately from the effect of the third level of B. This would be diagnosed differently depending on the software used with different combination of error messages, zero degrees of freedom for some effects in the ANOVA table, or a missing value for the statistic used for testing.

Table 1 - Example of factor disconnectedness.

Factor B Levels	Factor A Levels		
	1	2	3
1	x	x	
2	x	x	
3			x

In general, the variance between studies is large compared to the variance within studies, hence underlying the importance of including the study effect into the meta-analytical model. The study effect represents the combined effect of many factors that differ between studies, but which factors are not in the model because they either were not measured, or have been excluded from the model, or for which the functional form in the model is inadequately representing the true but unknown functional form (*e.g.*, the model assumes a linear relationship between the dependent and one independent continuous variable whereas the true relationship is nonlinear). In the absence of interactions between design variables (*e.g.*, studies) and the covariates (*e.g.*, all continuous independent variables of interest), parameter estimates for the covariates are unbiased, but the study effect adds uncertainty to future predictions (St-Pierre, 2001). The presence of significant interactions between studies and at least one covariate is more problematic since this indicates that the effect of the covariate is dependent on the study, implying that the effect of a factor is dependent on the levels of unidentified factors.

Steps in the meta-analytic process

There are several inherent steps to meta-analyses, the important ones being summarized in Figure 1. An important aspect of this type of analyses is the iteration process, which is under the control of the analyst. This circular pattern where prior steps are re-visited and refined is an important aspect of meta-analyses and contribute much to their heuristic characteristic.

Objectives of the study

Establishing a clear set of study objectives is a critical step that guides most ulterior decisions such as the database structure, data filtering, weighing of observations, and choice of statistical model. Objectives can cover a wide range of targets, ranging from preliminary analyses to identify potential factors affecting a system, thus serving an important role to the formulation of research hypotheses in future experiments, to the quantification of the effect of a nutritional factor such as a specific feed additive.

Data Entry

Results from prior research found in the literature must be entered in a database. The structure and coding of the database must include numerous variables identifying the experimental objectives. Hence numerous columns are added to code the objectives of each study. This coding is necessary so as to avoid the improper

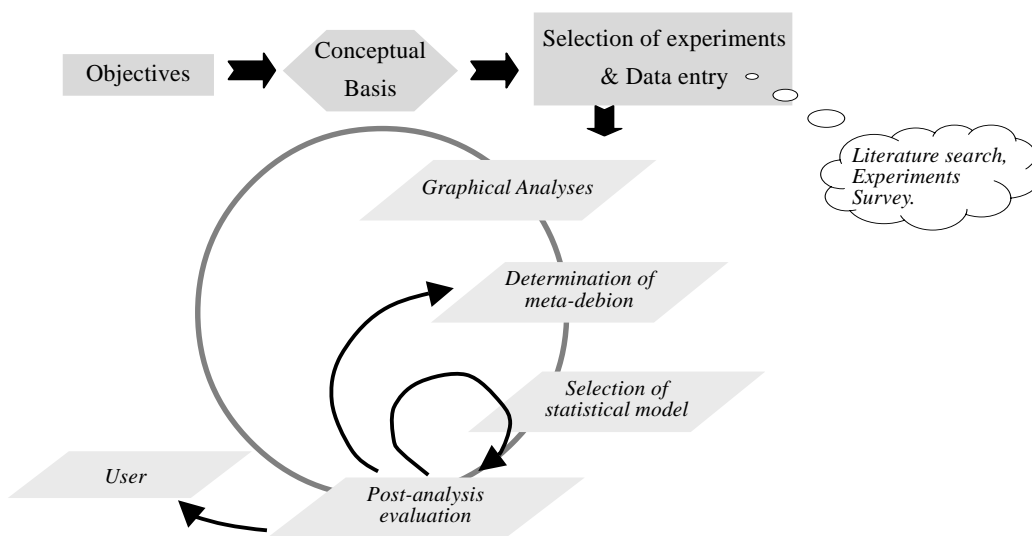


Figure 1 - Schematic representation of the meta-analytic process.

aggregation of results across studies with very different objectives. During this coding phase, the analyst may choose to transform a continuous variable to a discrete variable with n levels coded in a single column with levels of the discrete variable as entries, or in n columns with 0-1 entries to be used as dummy variables in the meta-analytic model. Different criteria can guide the selection of classes, such as equidistant classes, or classes with equal frequencies or probability of occurrence. The important point is that the sum of these descriptive columns must entirely characterize the objectives of all studies used.

Data Filtering

There are at least three steps necessary to effective data filtering. First, the analyst must ensure that the study under consideration is coherent with the objectives of the meta-analysis. That is, the meta-analytic objectives dictate that some variables must have been measured and reported. If, for example, the meta-analysis objective is to quantify the relationship between dietary NDF concentration and DM intake, then one must ensure that both NDF concentration and DM intake were measured and reported in all studies. The second step consists of a thorough and critical review of each publication under consideration, focusing on the detection of errors in the reporting of results. This underlines the importance of having a highly trained professional involved in this phase of the study. Only after publications have passed this *expert* quality filter should their results be entered in the database. Verification of data entries is then another essential component to the process. In this third step, it is important to ensure that a selected publication does not appear to be an outlier with respect to the characteristics and relations under consideration.

Preliminary Graphical Analyses

A thorough visual analysis of the data is an essential step to the meta-analytic process. During this phase, the analyst can form a global view regarding the coherence and heterogeneity of the data, as well as to the nature and relative importance of the inter-study and intra-study relationships of prospective variables taken two at a time.

Systematic graphical analyses should lead to specific hypotheses and initial selection of

alternate statistical models. Graphics can also help identifying observations that appear unique or even outliers. The general structure of relationships can also be identified, such as linear vs. nonlinear relationships as well as the presence of interactions. As an example, Figure 2 shows an intra-study curvilinear relationship between two variables in the presence of a significant inter-study effect (*i.e.*, different intercepts between studies). In this example, the inter-study effect associated with the X variable indicates the presence of a latent (hidden) variable that differed across studies. Another example is shown in Figure 3, which suggests the presence of a linear intra-study effect interacting with the study effect (*i.e.*, different regression slopes across studies). This may be due to a narrow range of the X variable in each experiment, or that again experiments differed with respect to a latent, interacting variable that was maintained constant or nearly constant within each experiment. This *visualization* phase of the data should always be taken as a preliminary step to the statistical analysis and not as conclusive evidence. The reason is that as the multi dimensions of the data are collapsed into two or possibly three dimension graphics, the unbalance that clearly is an inherent characteristic of meta-analytic data can lead to false visual relationships. This is because simple X-Y graphics do not correct the observations for the effects of all other variables that can affect Y.

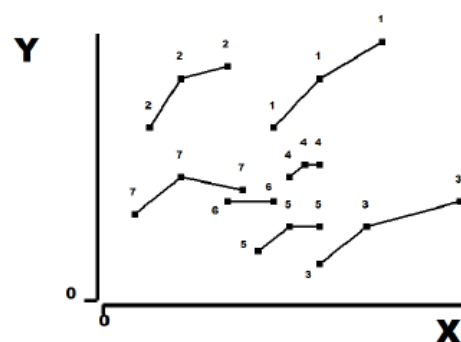


Figure 2 - Example of a curvilinear intra-study effect in the presence of a random study effect.

Graphical analyses should also be done in regards to the joint coverage of predictor variables, identifying their possible ranges, plausible ranges, and joint distributions, all being closely related to the inference range. Figure 4 shows the concepts involved when the dependent variable is plotted

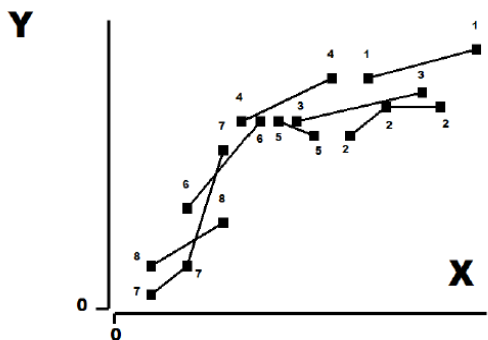


Figure 3 - Example of a linear intra-study effect in the presence of an interaction of study by predictor variable.

against one of the predictor variable. Similar graphics should be drawn to explore the relationships between predictor variables taken two at the time. In such graphics, the presence of any linear trends indicates correlations between predictor variables. Strong positive or negative correlations of predictor variables have two undesirable effects. First, they may induce near collinearity, implying that the effect of one predictor cannot be uniquely identified (*i.e.*, is nearly confounded with the effect of another predictor). In such instance, the statistical model can include only one of the two predictors at a time. Second, the range of a predictor X1 given a level of a second predictor X2 is considerably less than the unconditional range of predictor X1. In these instances, although the range of a predictor appears considerable in a univariate setting, its effective range is actually very much reduced in the multivariate space.

Figure 5 illustrates some of these concepts using an actual set of data on chewing activity in cattle and the NDF content of the diet. Visually, one concludes that the intra-study relationship between chewing activity and diet NDF is nonlinear, and that the intercept (*i.e.*, height of the curves) differed across studies. This observation can be formally tested using the statistical methods to be outlined later. In this example, the possible NDF range is 0 – 100%, whereas its plausible range is more likely between 20 and 60%. Interestingly, the figure illustrates that experimental measurements were frequently in the 20-40% and 45-55% ranges, leaving a gap with very few observations in the 40-45% range.

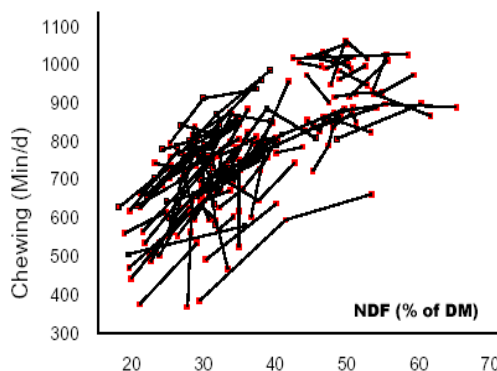


Figure 5 - Effect of dietary neutral detergent fiber (NDF) content on chewing activity in cattle. ¹Data are from 88 published experiments where the NDF content of the diet was the experimental treatment.

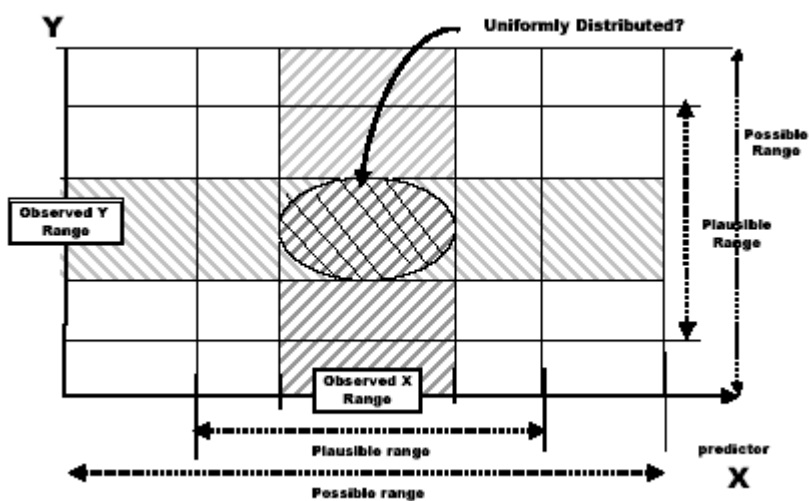


Figure 4 - Illustration of the concepts of possible range, plausible range, and actual range in a meta-analysis with a single predictor variable.

Study of the experimental meta-design

The meta-design is determined by the structure of the experiments for each of the predictor variables. To characterize the meta-design, numerous steps must take place before and after the statistical analyses. The specific steps depend on the number of predictor variables in the model.

One predictor variable.

- The experimental design used in each of the studies forming the database must be identified and coded, and their relative frequencies calculated. This information can be valuable during the interpretation of the results.
- Frequency plots (histograms) of the predictor variable can identify areas of focus of prior research. For example, Figure 6 shows the frequency distributions of NDF for the 517 treatment groups for the meta-analytic dataset used to draw Figure 5. Both figures indicate a substantial research effort towards diets containing 30-35% NDF, an area of dietary NDF density that borders the lower limit of recommended dietary NDF for lactating dairy cows (NRC, 2001). Because of the high frequency of observations in the 30-35% NDF range, the *a priori* expectation is that the effect of dietary NDF will be estimated most precisely in this range.
- One should also consider the intra and inter-study variances for the predictor variable. Small intra-study variances reduce the ability

of assessing the structural form of the relationship between the predictor and the dependent variable. Large intra-study variances but with only two levels of the predictor variable in all or most of the studies hides completely any potential nonlinear relationships. Figure 7 shows the intra-study standard deviation (S) of NDF as a function of the mean dietary NDF for experiments with two treatments, and for those with more than two treatments. The analyst can determine a minimum threshold of S to exclude experiments with little intra-study variation in the predictor variable. In instances where the study effect is considered random, this is not necessary and generally not desirable. In such instances the inclusion of studies with little variation in the predictor

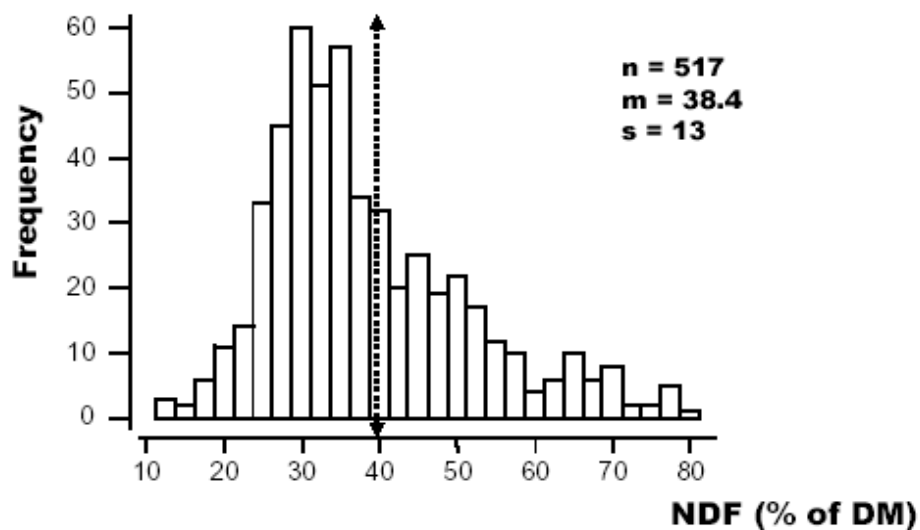


Figure 6 - Frequency plot of neutral detergent fiber (NDF) content of diets across experimental treatments.

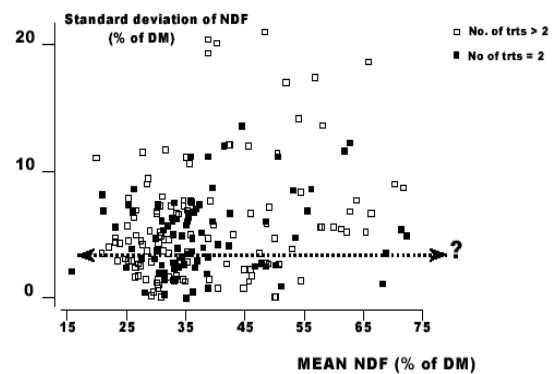


Figure 7 - Intra-study standard deviation of dietary neutral detergent fiber (NDF) and mean NDF of experiments designed to study the effects of NDF in cattle.

variable does little in determining the relationship between the predictor and the dependent variable, but adds observations and degrees of freedom to estimate the variance component associated with the study effect. When looking at a possible nonlinear intra-study relationship, it is intuitive to retain only those experiments with three or more levels of the predictor variables in the dataset. In this instance, intuition is incorrect. Experiments with two levels of the predictor variables add information on the study component (intercept) and the linear parameter of the model. In Figure 7, eliminating studies with less than three levels of the predictor variable would eliminate 40% of all the observations.

- Another important aspect at this stage of the analysis is to determine the significance of the study effect on the predictor variables. As explained previously in the context of Figures 4 and 5, one must be very careful in regards to the statistical model used to investigate the intra-study effect when there is an interaction between the predictor variable(s) and study. In these instances, the relationship between the predictor variable and the dependent variable is dependent on the study, which itself represents the sums of a great many factors such as measurement errors, systematic differences in the methods of measurements of the dependent variable across studies, and, more importantly, latent variables (hidden) not balanced across experiments. In those instances, the analyst must exert great caution in the interpretation of the results, especially regarding the applicability of these results.
- It is generally useful to calculate the leverage of each observation (Tomassone *et al.*, 1983). Traditionally, leverage values are calculated after the model is fitted to the data, but nothing prohibits the calculation of leverage values at an earlier stage because their calculations depend only on the design of the predictor variable in the model. For example, in the case of the simple linear regression with n observations, the leverage point for the i th observation is calculated as:

$$h_i = 1/n + (X_i - X_m) / \Sigma (X_i - X_m)^2 \quad [1]$$

where: h_i is the leverage value,

X_i is the value of the i^{th} predictor variable, and

X_m is the mean of all X_i .

Equation [1] clearly indicates that the leverage of an observation, *i.e.* its weight in the determination of the slope, grows with its distance from the mean of the predictor variable. The extension of the leverage point calculations to more than one predictor variables is straight forward (St-Pierre & Glamocic, 2000).

- In a final step, the analyst must graphically investigate the functional form of the relationship between the dependent variable and the predictor variable.

Two or more predictor variables

In the case of two or more predictor variables, the analyst must examine graphically and then statistically the inter and intra-study relationships between the predictor variables. Leverage values should be examined. With fixed models (all effects in the models are fixed with the exception of the error term), variance inflation factors (VIF) should be calculated for each predictor variables (St-Pierre & Glamocic, 2000). An equivalent statistic has not been proposed for mixed models (*e.g.*, when the study effect is random), but asymptotic theory would support the calculation of the VIFs for the fixed effect factors in cases where the total number of observations is large. The objective in this phase is to assess the degree of inter-dependence between the predictor variables. Because predictor variables in meta-analyses are never structured prior to their determination, they are always non-orthogonal and, hence, show variable degrees of inter-dependency. Collinearity determinations (VIF) assess one's ability to separate the effects of inter-dependent factors based on a given set of data. Collinearity is not model driven, but completely data driven.

Weighing of observations

Because meta-analytic data are extracted from the results of many experiments conducted under many different statistical designs and number of experimental units, the observations (treatment means) have a wide range of standard errors.

Intuition and classical statistical theory would indicate that observations should be subjected to some sort of weighing scheme. Systems used for weighing observations form two broad categories.

Weighing based on classical statistical theory

Under a general linear model where observations have heterogeneous but known variances, maximum likelihood parameters estimates are obtained by weighing each observation by the inverse of its variance. In the context of a meta-analysis where observations are least-squares (or population marginal) means, observations should be weighed by the inverse of the squares of their standard errors, which are the standard errors of each mean (SEM). Unfortunately, when such weights are used, the resulting measures of model errors (*i.e.*, standard error, standard error of predictions, etc.) are no longer expressed in the original scale of the data. To maintain the expressions of dispersion in the original scale of the measurements, St-Pierre (2001) suggested dividing each weight by the mean of all weights, and to use the resulting values as weighing factors in the analysis. Under this procedure, the average weight used is algebraically equal to 1.0, thus resulting in expressions of dispersion that are in the same scale as the original data.

Weighing based on other criteria

Other weighing criteria have been suggested for the weighing of observations, such as the power of an experiment to detect an effect of a size defined *a priori*, the duration of an experiment, etc. The weighing scheme can actually be based on an expert assessment, partially subjective, of the overall quality (precision) of the data. The opinion of more than one expert may be useful in this context. From a Bayesian statistical paradigm, the use of subjective information for decision-making is perfectly coherent and acceptable, as subjective probabilities are often used to establish prior distributions in Bayesian decision theory (De Groot, 1970). Traditional scientific objectiveness, however, may restrict the use of this weighing scheme in scientific publications.

Predictably, the importance of weighing observations decreases with the number of observations used in the analysis, especially if the

observations that would receive a small weight have relatively small leverage values. An example of this is shown in Figure 8.

Whether the analyst should weight the observations based on the SEM for each individual treatments or the pooled SEM from the studies is open for debate. There are many reasons why the SEM of each treatment within a study can be different. First, the original observations themselves could have been homoscedastic (homogeneous variance) but the least-squares means would have different SEM due to unequal frequencies (*e.g.*, missing data). In such case, it is clear that the weight should be based on the SEM of each treatment. Second, the treatments may have induced heteroscedasticity, meaning that the original observations did not have equal variances across sub-classes. In such instances, the original authors should have conducted a test to assess the usual homoscedasticity assumption in linear models. The problem is that a lack of significance (*i.e.*, $P > 0.05$) when testing the homogeneity assumption does not prove homoscedasticity, but only that the null hypothesis (homogeneous variance) cannot be rejected at a $P < 0.05$. In a meta-analytic setting, the analyst may deem the means with larger apparent variance to be less credible and reduce the weigh of these observations in the analysis. Unfortunately, most publications lack the information necessary to this option.

Among the more subjective criteria available for weighing is the quality of the experimental design used in the original study in regard to the meta-analytic objective.

Experimental designs have various trade-offs

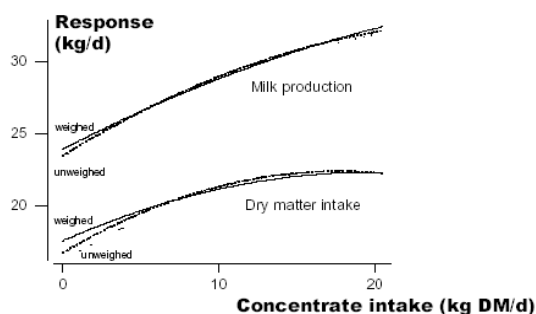


Figure 8 - Estimated response of lactating dairy cows to concentrate intake using weighed and unweighed observations.

due to their underlying assumptions. For example, the Latin square is often used in instances where animal units are relatively expensive, such as in metabolic studies. The double orthogonal blocking used to construct Latin square designs can remove a lot of variation from the residual error.

Thus very few animals can be used compared to a completely randomized design for an equal power of detecting treatment effect. The downside, however, is that the periods are generally relatively short to reduce the likelihood of a period by treatment interaction (animals in different physiological status across time periods), thus reducing the magnitude of the treatment effects on certain traits, such as production and intake for example. In those instances, the analyst should legitimately weigh down observations from experiments whose designs limited the expression of the treatment effects.

Statistical models

The independent variables can be either discrete or continuous. With binary data (healthy/sick, for example), generalized linear models (GLM) based on the logit or probit link functions are generally recommended (Agresti, 2002). Because of advances in computational power, the GLM has been extended to include random effects in what is called the generalized linear mixed model (GLMM). In its version 9, the SAS system includes a beta release of the GLIMMX procedure to fit these complex models.

In nutrition, however, the large majority of the dependent variables subjected to meta-analyses are continuous, and their analyses are treated at length in the remainder of this paper.

St-Pierre (2001) made a compelling argument to include the study effect in all meta-analytic models. Because of the severe imbalance in most databases used for meta-analyses, the exclusion of the study effect in the model leads to biased parameter estimates of the effects of other factors under investigation, and severe biases in variance estimates. In general, the study effect should be considered random because it represents the sum of the effects of a great many factors, all with relatively small effects on the dependent variable. Statistical theory indicates that these effects would be close to Gaussian (normal), thus much better estimated if treated as random effects.

Practical recommendations regarding the selection of the type of effect for the studies are presented in Table 2. In short, the choice depends on the size of the conceptual population, and the sample size (the number of studies in the meta-analysis).

The ultimate (and correct) meta-analysis would be one where all the primary (raw) data used to perform the analyses in each of the selected publications were available to the analyst. In such instance, a large segmented model that includes all the design effects of the original studies (*e.g.*, the columns and rows effects in Latin squares) plus the effects to be investigated by the meta-analysis could be fitted by least-squares or maximum likelihood methods. Although computationally complex, such huge meta-analytic models should be no more difficult to solve than the large models used by geneticists to estimate the breeding values of animals using very large national databases of production records. Raw data availability should not be an issue in instances where meta-analyses are conducted with the purpose of summarizing research at a given research center. This, however, is very infrequent, and meta-analyses are almost always conducted using observations that are themselves summaries of prior experiments (*i.e.*, treatment means). It seems evident that a meta-analysis conducted on summary statistics should lead to the same results as a meta-analysis conducted on the raw data, which itself would have to include a study effect because the design effects are necessarily nested within studies (*e.g.*, cow 1 in the Latin square of study 1 is different than the cow 1 of study 2), which itself would be considered random. Thus, analytical consistency dictates the inclusion of the study effect in the model, generally as a random effect. The study effect will be considered random in the remainder of this paper, with the understanding that under certain conditions explained previously it should be considered a fixed effect factor.

Beside the study effect, meta-analytic models include one or more predictor variables that are either discrete, or continuous. For clarity, we will initially treat the case of each variable type separately, understanding that a model can easily include a mixture of variables of both types, as we describe later.

Model with discrete predictor variable(s)

A linear mixed model easily models this situation as follows:

$$Y_{ijk} = \mu + S_i + \tau_j + S_{ij} + e_{ijk} \quad [4]$$

where:

Y_{ijk} = the dependent variable,

μ = overall mean,

S_i = the random effect of the i^{th} study, assumed $\sim_{\text{iid}} N(0, \sigma_s^2)$,

τ_j = the fixed effect of the j^{th} level of factor τ ,

S_{ij} = the random interaction between the i^{th} study and the j^{th} level of factor τ , assumed $\sim_{\text{iid}} N(0, \sigma_{s\tau}^2)$, and

e_{ijk} = the residual errors, assumed $\sim_{\text{iid}} N(0, \sigma_e^2)$.

e_{ijk} , S_{ij} and S_i are assumed to be independent random variables.

For simplicity reasons, model [4] is written without weighing the observations. The weights would appear as multiplicative factors of the diagonal elements of the error variance-covariance matrix (Draper & Smith, 1998). Model [4] corresponds to an incomplete, unbalanced randomized block design with interactions in classic experimental research. The following SAS statements would solve this model:

```
PROC MIXED DATA=Mydata CL COVTEST;
  CLASSES study tau;
  MODEL Y = tau;
  RANDOM study study*tau;
  LSMEANS tau;
RUN;
```

Standard tests of significance on the effect of τ are easily conducted and least-squares means can be separated using an appropriate mean separation procedure. Although it may be tempting to remove the study effect from the model in instances where it is not significant (also called pooling of effects), this practice can lead to biased probability estimations (*i.e.*, final tests on fixed effects are conditional on tests for random effects) and is not recommended. This is because not being able to reject the null hypothesis of no study effect (*i.e.*, variance due to study is not significantly different from zero) is a very different proposition

than proving that the effect of study is negligible. At the very least, the probability threshold for significance of study should be much larger than the traditional $P = 0.05$. Ideally, the analyst should state before the analysis is performed what size of estimated variance due to study should be considered negligible, such as $\sigma_s^2 < 0.1 \sigma_e^2$.

Model with continuous predictor variable(s)

A linear mixed model is used:

$$Y_{ij} = B_0 + S_i + B_1 X_{ij} + b_i X_{ij} + e_{ij} \quad [6]$$

where:

Y_{ij} = the dependent variable,

B_0 = overall (inter-study) intercept (a fixed effect equivalent to τ in [4]),

S_i = the random effect of the i^{th} study, assumed $\sim_{\text{iid}} N(0, \sigma_s^2)$,

B_1 = the overall regression coefficient of Y on X (a fixed effect),

X_{ij} = value of the continuous predictor variable,

b_i = random effect of study on the regression coefficient of Y on X , assumed $\sim_{\text{iid}} N(0, \sigma_b^2)$, and

e_{ij} = the residual errors, assumed $\sim_{\text{iid}} N(0, \sigma_e^2)$.

e_{ij} , b_i and S_i are assumed to be independent random variables.

The following SAS statements solve this model:

```
PROC MIXED DATA=Mydata CL COVTEST;
  CLASSES study;
  MODEL Y = X/SOLUTION;
  RANDOM study study*X;
RUN;
```

Using a simple Monte Carlo simulation, St-Pierre (2001) demonstrated the application of this model to a synthetic dataset, showing the power of this approach, and the interpretation of the estimated parameters.

Model with both discrete and continuous predictor variable(s)

Statistically, this model is a simple combination of [4] and [6] as follows:

$$Y_{ijk} = \mu + S_i + \tau_j + S_{ij} + B_1 X_{ij} + b_i X_{ij} + B_j X_{ij}$$

$$+ e_{ijk} \quad [8]$$

where:

B_j = the effect of the j^{th} level of the discrete factor τ on the regression coefficient (a fixed effect).

The following SAS statements would be used to solve this model:

```
PROC MIXED DATA=Mydata CL COVTEST;
  CLASSES study tau;
  MODEL Y = tau X tau*X;      [9]
  RANDOM study study*tau study*X;
  LSMEANS tau;
RUN;
```

In theory, [8] is solvable, but the large number of variance components and interaction terms that must be estimated, in combination with the imbalance in the data makes it often numerically intractable. In such instances, at least one of the two random interactions must be removed from the model.

In [4], [6], and [8], the analyst secretly wishes for the interactions between study and the predictor variables to be highly nonsignificant. Recall that the study effect represents an aggregation of the effects of many uncontrollable and unknown factors that differed between studies. A significant study \times τ interaction in [4] implies that the effect of τ (the intercept) is dependent on the study, hence of factors that are unaccounted for. Similarly, a significant interaction of study by X in [6] (the b_j terms) indicates that the slope of the linear relationship of Y on X is dependent on the study, hence of unidentified factors. In such situation, the analysis produces a model that can explain very well the observations, but whose predictions of future outcomes are generally not precise because the actual realization of a future study effect is unknown.

The maximum likelihood predictor of a future observation is computed by setting the study and the interaction of study with the fixed effect factors to their mean effect values of zero (McCulloch & Searle, 2001), but the standard error of this prediction is very much amplified by the uncertainty regarding the realized effect of the future study.

When the study effect and its interaction with

fixed effect is correctly viewed as an aggregation of many factors not included in the model, but that differed across studies, the desirability of including as many fixed factors in the model as can be uniquely identified from the data becomes obvious. In essence, the fixed effects should ultimately make the study effect and its interactions with fixed effects predictors small and negligible. In such instances, the resulting model should have wide forecasting applicability. Imagine for instance that much of the study effect on body weight gain of animal is in fact due to large difference in the initial body weight across studies. In such instances, the inclusion of initial body weight as a covariate would remove much of the study effect, and the diet effects (as continuous or discrete variables) would be estimated without biases, with a wide range of applicability (*i.e.*, a future prediction would require a measurement of initial body weight as well as measurements of the other predictor variables).

Whether one chooses [4] or [6] as a meta-model is somewhat arbitrary when the predictor variable has an inherent scale (*i.e.*, is a measured number). The assumptions regarding the relationship between Y and the predictor variable are, however, very different between the two models. In [4], the model does not assume any functional form for the relationship. In [6], the model explicitly assumes a linear relationship between the dependent and the predictor variables. Different methods can be used to determine whether the relationship has a linear or nonlinear structure.

- The first method consists in classifying observations into five sub-classes based on the quintiles for the predictor variable, and performing the analysis according to [4] with five discrete levels of the predictor variable. Although the selection of five sub-classes is somewhat arbitrary, there are substantive references in the statistical literature indicating that this number of levels generally works well (Cochran, 1968; Rubin, 1997). A visual inspection of the five least-squares means, or the partitioning of the four degrees of freedom associated with the five levels of the discrete variable into singular orthogonal polynomial contrasts can rapidly identify an adequate functional form to use for modeling the Y-X relationship.

- The second method can be directly applied to the data, or can be a second step that follows the identification of an adequate degree for a polynomial function. Model [6] is augmented with the square (and possibly higher order terms) of the predictor variable. In the MIXED procedure of SAS, this can be done simply by adding an X^2 term to the model statement. It is important to understand that in the context of a linear (mixed or not) model, the matrix representation of the model and the solution procedure used are no different when X and X^2 are in the model compared to a situation where two different continuous variables (say X and Z) are included in the model. The problem, however, is that X and X^2 are implicitly dependent; after all, there is an algebraic function relating the two. This dependence can result in a large correlation between the two variables, thus leading to possible problems of collinearity.
- A third method can be used in more complex situations where the degree of the polynomial exceeds two, or the form of the relationship is sigmoid, for example. The relationship can be modeled as successive linear segments, an approach conceptually close to the first method explained previously. Martin and Sauvant (2002) used this method to study the variation in the shape of the lactation curves of cows subjected to various concentrate supplementation strategies, using the model of Grossman & Koops (1988) as its fundamental basis. Using this approach, lactation curves were summarized by a vector of 9 parameter estimates, which estimates could be compared across supplementation strategies.

In [8], the interest may be in the effect of the discrete variable τ after adjusting for the effect of a continuous variable X as in a traditional covariate analysis, or the interest may be inverse, *i.e.*, the interest is in the effect of the continuous variable after adjusting for the effect of the discrete variable. The meta-analyses of Firkins *et al.* (2001) provide examples of both situations. In one instance, the effect of grain processing (a discrete variable) on milk fat content was being

investigated while correcting for the effect of dry matter intake (DMI, a continuous variable). In this instance, the interest was in determining the effect of the discrete variable. In another instance, the effects of various dietary factors such as dietary NDF, DMI and proportion of forage in the diet (all continuous variables) on starch and NDF digestibility, and microbial N synthesis were investigated, while correcting for the effect of the method of grain processing. In this case, the interest was much more towards the effects of the continuous variables exempt from possible biases due to different grain processing methods across experiments.

Accounting for interfering factors

Differences in experimental conditions between studies can affect the treatment response. The nature of these conditions can be represented by quantitative or qualitative variables. In the first instance the variable and possibly its interaction with other factors can be added to the model if there are sufficient degrees of freedom. The magnitude of treatment response is sometimes dependent on the observed value in the control group. For example, Figure 9 shows the milk fat response in lactating cows to dietary buffer supplementation as a function of the milk fat of the control group (Meschy *et al.*, 2004). The response was small or non-existent when the milk fat of control cows was near 40 g/L, but increased markedly when the control cows had low milk fat, possibly reflecting a higher likelihood of sub-clinical rumen acidosis in these instances.

The presence of a study by predictor variable interaction can indicate a nonlinear relationship and the need for a higher degree of polynomial in the model. Applying model [6] with the addition of a square term for the X_{ij} to the data shown in Figure 5 results in a relatively good quantification of the relationship between chewing time and dietary NDF, as shown in Figure 10. In this type of plot, it is important to adjust the observations for the study effect, or the regression may appear to poorly fit the data because of the many hidden dimensions represented by the studies (St-Pierre, 2001).

In instances where the interacting factor is discrete, the examination of the sub-classes least-squares means can clarify the nature of the

interaction. For example, the effect of a dietary treatment may be dependent on the physiological status of the animals used in the study. This physiological status can be coded using multiple dummy variables, as explained previously.

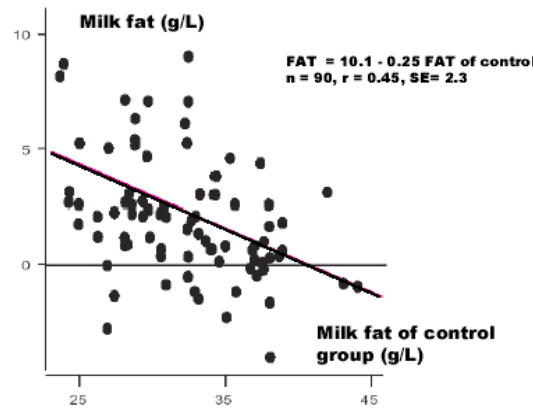


Figure 9 - Response in milk fat content (FAT) to dietary buffer supplementation.

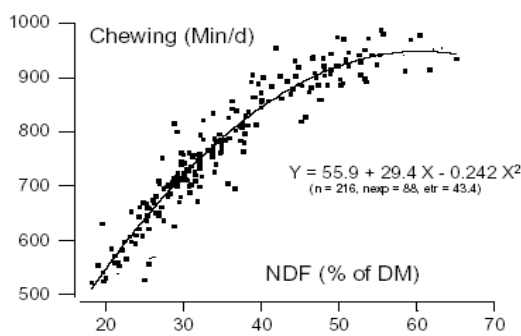


Figure 10 - Effect of dietary neutral detergent fiber (NDF) content on chewing activity in cattle. ¹Data are from published experiments where the NDF content of the diet was the experimental treatment. Observations were adjusted for the study effect before being plotted, as suggested by St-Pierre (2001).

Post-optimization analyses

As when fitting conventional statistical models, numerous analyses should follow the fitting of a meta-analytic model. These analyses are used to assess the assumptions underlying the model, and to determine whether additional meta-analytic models should be investigated.

Structure of Residual Variation

In [4], [6], and [8], the residuals (errors) are assumed independent, and identically distributed from a normal population with a mean of zero and

a variance σ^2e . The normality assumption can be tested using a standard Chi^2 test, or a Shapiro-Wilkes test, both available in the UNIVARIATE procedure of the SAS system. Residuals can also be expressed as Studentized residuals, with absolute values exceeding 3 being suspected outliers (Tomassone *et al.*, 1983). In meta-analysis, the removal of a suspected outlier observation should be done only with extreme caution. This is because the observations in a meta-analysis are the calculated outcomes (least-squares means) of models and experiments that should themselves be nearly free of the influence of outliers. Thus, meta-analytic outliers are much more likely indicative of a faulty model than of a defective observation. In addition, the removal of one treatment mean as an observation in a meta-analysis might be removing all the variation in the predictor variable for the experiment in question, thus making the value of the experiment in a meta-analytic setting nearly worthless. In addition, the analyst should examine for possible intra and inter-study relationships between the residuals and the predictor variables.

Structure of study variation

When a model of the type described in [4], [6], or [8] is being fitted, it is possible to examine each study on the basis of its own residuals. For example, Figure 11 shows the distribution of the residual standard errors for the different studies used in the meta-analysis of chewing time in cattle (Figures 5 and 10). Predictably, the distribution is asymmetrical and follows the law of Raleigh for the standard errors, while variances have a Chi^2 distribution. Studies with unusual standard errors, say those with a Chi^2 probability exceeding 0.999 could be candidates for exclusion from the analysis. Alternatively, one could consider using the inverse of the estimated standard errors as weights to be attached to the observation before re-iterating the meta-analysis.

Other calculations such as leverage values, Cook's distances, and other statistics can be used to determine the influence of each observation on the parameter estimates (Tomassone *et al.*, 1983).

Table 2 - Guidelines to establish whether an effect should be considered fixed or random in a meta-analytic model¹.

	Population	Experiment	Effect of t in the model
Case 1	T is small ²	$t \cong T$	Fixed effect
Case 2	T is large	$t \ll T$	Random effect
Case 3	T is large	$t \cong T$	Should be fixed but random works better
Case 4	T is large	$t \ll T$, and t is very small	Should be random but fixed may work better (<i>i.e.</i> , variance components can be poorly estimated)

¹Adapted from Milliken (1999) for mixed models. ²T represents the number of studies in the population (conceptual), and t is the number of studies in the meta-analysis.

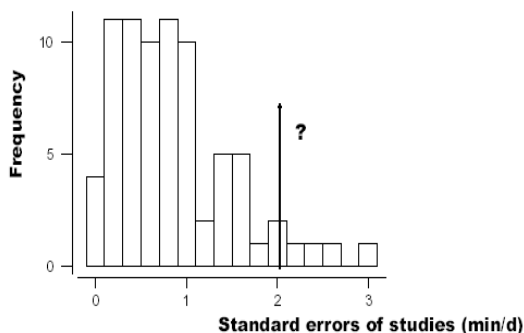


Figure 11 - Frequency distribution of the residual standard errors of the studies used in the meta-analysis of chewing activity in cattle.

Conclusion

Meta-analyses produce empirical models. They are invaluable for the synthesis of data that at first may appear scattered without much pattern. The meta-analytic process is heuristic and implicitly allows returning to prior steps. Extensive graphical analyses must be performed prior to the parameterization of a statistical model to gain a visual understanding of the data structure as well as to validate data entries.

The increased frequency of meta-analyses published in the scientific literature coupled with scarce funding for research should create an additional need for scientific journal to ensure that published articles provide sufficient information to be used in a subsequent meta-analysis. There may be a time when original data from published articles will be available via the web in a standardized format, a current practice in DNA sequencing research.

Lastly, new meta-analytic methods should assist the expansion of mechanistic modeling efforts of complex biological systems by providing conceptual models as well as a structured process for their external evaluation.

Acknowledgment

The author is thankful to Daniel Sauvant, Philippe Schmidely, and J. J. Daudin for their contribution to this paper. Much of the material presented in this paper was shared by these colleagues during a visit by the author to the Institut National Agronomique (INRA INAPG) Paris-Grignon in 2004. Additionally, parts of this paper appeared previously in an internal INRA discussion paper (INRA Productions Animales) written by Professor Sauvant.

Literature citad

- AGRESTI, A. **Categorical Data Analysis**, 2 ed. New York: John Wiley & Sons, 2002. 710p.
- COCHRAN, W.G. The effectiveness of adjustment by subclassification in removing bias in observational studies. **Biometrics**, v.24, p.295-313, 1968.
- DE GROOT, M.H. **Optimal Statistical Decision**. New York: McGraw-Hill, 1970. 489p.
- DRAPER, N.R.; SMITH, H. **Applied Regression Analysis**. New York: John Wiley & Sons, Inc., 1998. 706 p.
- EUGÈNE, M.; ARCHIMÈDE, H.; SAUVANT D. Quantitative meta-analysis on the effects of defaunation of the rumen on growth, intake and digestion in ruminants. **Livestock Production Science**, v.85, p.81-91, 2004.
- FIRKINS, J.L.; EASTRIDGE, M.L.; ST-PIERRE, N.R. *et al.* Invited: Effects of grain variability and processing on starch utilization by lactating dairy cattle. **Journal of Animal Science**, v.79(E. Suppl.), p.E218-E238, 2001.
- GLASS, G.V. Primary, secondary and meta-analysis of research. **Education Research**, v.5, p.3-8, 1976.
- GROSSMAN, M.; KOOPS, W.J. Multiphasic analysis of lactation curves in dairy cattle. **Journal of Dairy Science**, v.71, p.1598-1608, 1988.
- MANTEL, N.; HAENSZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. **Journal of the National Cancer Institute**, v.22, p.719-748, 1959.
- MARTIN, O.; SAUVANT, D. Meta-analysis of input/output kinetics in lactating dairy cows. **Journal of Dairy Science**, v.85, p.3363-3381, 2002.
- McCULLOCH, C.E.; SEARLE, S.R. **Generalized, Linear, and Mixed Models**. New York: John Wiley & Sons, Inc., 2001. 325p.
- MESCHY, F.; BRAVO, D.; SAUVANT, D. Analyse quantitative des réponses des vaches laitières à l'apport de substances tampon. **INRA Production Animale**, v.17, n.1, p.11-18, 2004.
- MILLIKEN, G. A. Mixed Models. Seminar to the Cleveland

- Chapter of the American Statistical Association, 1999. 120 p.
- National Research Council. **Nutrient Requirements of Dairy Cattle**. 7ed. Washington, D.C.: National Academic Press, 2001. 381p.
- OFFNER, A.; BACH, A.; SAUVANT, D. Quantitative review of *in situ* starch degradation in the rumen. **Animal Feed Science and Technology**, v.106, p.81-93, 2003.
- OFFNER, A.; SAUVANT, D. Prediction of *in vivo* starch digestion in cattle from *in situ* data. **Animal Feed Science and Technology**, v.111, p.41-56, 2004.
- OLDICK, B.S.; FIRKINS, J.L.; ST-PIERRE, N.R. Estimation of microbial nitrogen flow to the duodenum of cattle based on dry matter intake and diet composition. **Journal of Dairy Science**, v.82, p.1497-1511, 1999.
- RUBIN, D.B. Estimating causal effects from large data sets using propensity scores. **Annals of Internal Medicine**, v.127, p.757-763, 1997.
- SAUVANT, D.; MARTIN O. Empirical modelling through meta-analysis vs mechanistic modelling. In: KEBREAB, E. (Ed.) **Nutrient Digestion and Utilization in Farm Animals: Modelling approaches**. London: CAB International, 2004.
- SAUVANT, D.; MERTENS, D. Relationship between fermentation and liquid outflow rate in the rumen. **Reproduction Nutrition Development**, v.40, p.206-207, 2000.
- ST-PIERRE, N.R. Invited review: integrating quantitative findings from multiple studies using mixed model methodology. **Journal of Dairy Science**, v.84, p.741-755, 2001.
- ST-PIERRE, N.R. Reassessment of biases in predicted nitrogen flows to the duodenum by NRC 2001. **Journal of Dairy Science**, v.86, p.344-350, 2003.
- ST-PIERRE, N.R.; GLAMOCIC, D. Estimating unit costs of nutrients from market prices of feedstuffs. **Journal of Dairy Science**, v.83, p.1402-1411, 2000.
- TOMASSONE R.; LESQUOY, E.; MILLIER C. La régression: nouveau regard sur une ancienneméthode statistique. In: MASSON (Ed.) **Actualités Scientifiques et Agronomiques de l'INRA**. Paris:INRA, 1983.