

## Perspective: are animal scientists forgetting the scientific method and the essential role of statistics?

J. L. Black<sup>A,D</sup>, S. Diffey<sup>B</sup> and S. G. Nielsen<sup>C</sup>

<sup>A</sup>John L Black Consulting, PO Box 4021, Warrimoo, NSW 2774, Australia.

<sup>B</sup>Centre for Crop and Disease Management, Curtin University, Bentley, NSW 6102, Australia.

<sup>C</sup>Research Office, Charles Sturt University, Wagga Wagga, NSW 2650, Australia.

<sup>D</sup>Corresponding author. Email: [jblack@pnc.com.au](mailto:jblack@pnc.com.au)

**Abstract.** Animal scientists and their funding organisations need to ensure investment in research is maximised by strict adherence to the scientific method and the rigorous design and analysis of experiments. Statisticians should be considered as equals in the research process, engaged from the beginning of research projects and appropriately funded. The importance of experimental design that accounts for factors affecting the primary experiment measurement is illustrated in two examples. One shows how failure to involve a statistician at the beginning of a project resulted in considerable waste of resources. Subsequent engagement of professional statisticians with rigorous experimental design and analysis led to greatly increased precision in the standard error of an estimate for the digestible energy content of cereal grains for pigs from  $\pm 0.35$  MJ/kg to  $\pm 0.16$  MJ/kg. The other example shows the effect of the percentage of diets replicated during pelleting and of the total number of pigs required in the experiment on the *P*-values associated with detecting a pairwise difference between two grains differing in digestible energy content by 0.33 MJ/kg. Decisions based on these relationships have animal welfare and resource allocation implications.

**Additional keywords:** experimental design, measurement accuracy, treatment replications, digestible energy, pigs.

Received 5 June 2015, accepted 21 August 2015, published online 22 January 2016

### Introduction

As a Research Management Consultant overseeing multiple research projects for many organisations, one of us (JLB) is concerned by the apparent lack of scientific rigour in the development and analysis of experiments within sectors of animal science. Near identical repetition of earlier research occurs, hypotheses are often not clearly stated and rigorous experiment design and analysis is frequently lacking. This lack of strict adherence to the scientific method results in poorer outcomes in terms of understanding the principles of animal science and reduces returns for research investors, which in Australia are often grower-funded research organisations.

The role of statistics in the scientific method is critical, yet it appears that many scientists regard statisticians as ‘technicians’ rather than as equals in the research process. Frequently, statisticians are engaged after an experiment has been completed. R. A. Fisher, the originator of many statistical principles and of the basic statistical methodology on experimentation, wrote: ‘To consult the statistician after an experiment is finished is often merely to ask him to conduct a *post mortem* examination. He can perhaps say what the experiment died of’ (Fisher 1938). Similarly, many funding organisations appear not to appreciate the essential role of statisticians in research and fail to provide funds for proper experimental design, analysis and reporting. This Journal Perspective article

reiterates the steps in and importance of strict adherence to the scientific method and highlights the critical role of statisticians within the research process.

### The scientific method

Although there may be some argument about the precise steps within the scientific method, the general format is covered by the following nine tasks.

#### 1. Review literature and other relevant information

A thorough review of published work is critical to ensure a full understanding of all relevant research that has been conducted within a subject area. A thorough examination of previous research should ensure that earlier research is not ‘reinvented’ and that new concepts about the way a system functions can be formulated. These new concepts can frequently be tested with reanalysis or reorganisation of existing results within the literature to allow further development of the ideas before undertaking the critical experiments. The enormity of existing information should never be under-estimated and considerable effort is often required to formulate sound new ideas for research. A thorough examination of the literature is required to identify factors that may influence the concepts being developed as these may also need to be measured within any new experiment. The literature review can also reveal issues that may cause difficulty

for the proposed research program and provide guidance for the statistical design, analysis and reporting of the information being collected.

## 2. Establish a hypothesis

Hypotheses are formed from examination of previous research findings, sometimes by thorough reanalysis of existing knowledge, including possible anomalous findings in earlier research, or sometimes by serendipitous observations. From whichever process an idea for research arises, development of a clear hypothesis is essential before an experiment is undertaken. There are some scientists who believe there is value in the 'look-and-see' type approach. Although this can sometimes be helpful in the process of scientific discovery, we believe strongly that the best scientific outcomes result from the documentation of an objective hypothesis that can be tested by experimentation. Outcomes from investigations into largely unknown areas of knowledge for which the 'look-and-see' approach is sometimes advocated, we believe will be improved if based on a formulated hypothesis.

## 3. Design an experiment to test the hypothesis

There are two components in designing an experiment. First is the experimental protocol, which sets out the treatments and methods needed to test the hypothesis. Second is the actual design of the experiment, which should account for spatial and/or temporal variation in facilities and experiment duration. Typically 'blocking' is used to accommodate these sources of variation. However, modelling additional sources of variation at the analysis stage may also be appropriate. Factors that should be considered when undertaking the design include:

- What are the populations of interest?
- What treatments are required to test the hypothesis?
- What is the unit on which the measurement will be recorded?
- What measurements are to be made and how will they be made?
- What is the magnitude of differences between treatments needed to test the hypothesis?
- What is the variance of the measurement and what factors affect the values being measured?
- Which of these factors that affect the primary measurement should also be measured?
- What is the likely estimated residual error variance associated with the measured variable after accounting for factors known to affect the measurement?
- How many replicates are therefore required to show a selected level of significance (e.g.  $P < 0.05$ ) for the predetermined difference identified to validate/falsify the hypothesis? This requires a 'power' analysis and includes consideration of Type I errors (false positives) and Type II errors (false negatives).
- What design meets the essential criteria and fits best within physical, animal welfare and financial constraints?

Statisticians must be involved early in the decision-making process and are essential for designing the experiment. It is critical that spatial variation in facilities and temporal variations in measurements are effectively accommodated within any design. The exact number of each type of sample that is to be collected will be identified in the experimental design and

labels for these samples can be printed before the experiment is conducted.

## 4. Conduct the experiment

The experiment is conducted in strict adherence to the experimental protocol and design. Attention to detail in following all procedures including sampling protocols, measurement techniques and recording is essential to reduce the level of residual error associated with measurements being made in the experiment. When components of the design are disrupted because of unforeseen events such as a sick animal, the statistician should be consulted to ensure that the missing value can be most effectively replaced (or ignored) and the analytical power of the experiment maintained or compromised least. If the statistician has been involved from the start of the project, these decisions are often made quickly with no disruption to the experiment. In circumstances where it is not possible to consult a statistician, detailed documentation of observations and actions is essential.

## 5. Statistically analyse the results

Analysis of experimental results is straight forward when the experiment has been properly designed. Statistical models can be fitted that account for all measured variables and where appropriate their interactions. The residual error variance for testing treatment differences will be minimised when factors affecting measured values are accounted for within the statistical model. The magnitude of the differences between treatments that are statistically significant will then be reduced and the power of the experiment maximised. Statistical analysis of poorly designed experiments is often difficult and, in some cases, impossible.

## 6. Interpret the results

Rigorous experimental design and statistical analysis aids interpretation of the results in relation to the initial hypothesis. Effective use of statistical graphics can be a powerful aid to exploring possible new relationships or alternative explanations, leading to further concept development and experimentation. Interpretation of results can also be enhanced by incorporating concepts about operation of a system into mathematical equations within simulation models (Baldwin 1995). This approach enables theories and concepts to be evaluated quantitatively and provides a rigorous method for extending ideas about the functioning of a system.

## 7. Draw conclusions

Succinct conclusions that quantify outcomes and uncertainties from the findings are then drawn from the interpreted results for the information of others.

## 8. Prepare a manuscript

The experiment is not completed until it has been described and published for others to review, criticise and replicate if desired. Consideration should be given to publishing or reporting – in a retrievable form – non-significant results because the literature can be biased towards a minority of significant results, which are unrepresentative of the total research conducted on a topic

(Levine 2013). There is an increasing trend to submit electronically all results from experiments to journals or specified databases for reanalysis by others and this practice should be encouraged. Preparation of the manuscript takes time and the cost should be included in contracts with funding organisations.

## 9. Have the manuscript peer reviewed

Peer review of the final manuscript is an essential step in the path to publication. The best process for peer review commences with critical assessment by knowledgeable colleagues or associates before the formal process conducted by editors of scientific journals.

## An example of the essential role of statisticians

A large, integrated Australian research program, the Premium Grains for Livestock Program, was established to develop a rational basis for trading grains for livestock based on development of near-infrared spectroscopy (NIR) calibrations (Black 2008). A major component of the project was to predict the energy value of cereal grains for different livestock types. The project is continuing for pigs and broiler chickens. Over 3700 cereal grains (wheat, barley, oats, triticale, sorghum and maize) with a wide range in chemical and physical characteristics have been collected from germplasm archives, plant breeders, farmers, and selected because of drought, frost damage or pre-harvest germination. Approximately 400 of these grains have been fed to sheep, cattle, pigs, broilers and layers and energy disappearance along the digestive tract was measured. There is one overall experiment for each animal type. However, this experiment comprises many sub-experiments spread over time and, for some livestock types, over sites. For example, 13 sub-experiments beginning in 1997 and concluding in 2016 will be undertaken to develop and evaluate the NIR calibrations for pigs.

Each sub-experiment for pigs involves the feeding of 20–40 grains. Each grain represents a constant percentage of the diet, which is ground and cold-pelleted. Each pig is fed many grains over time with 4–5 experimental periods per sub-experiment. The accuracy of the NIR calibration is dictated by the accuracy of the estimates of digestible energy content of the grains from the experiments. When the research program was initiated, no funds were allocated for a statistician. The experiments were designed by the research personnel. Only one grain was used for connectivity between sub-experiments and there was little rigour in the design for spatial variability within pig feeding facilities or for temporal variation when feeding pigs over time. When a statistician was eventually asked to analyse the experiments, the variation in estimates of digestible energy of the grains was so large that the data proved unsuitable for developing NIR calibrations. Results from the first 3 years of this research were discarded for the purposes of NIR calibration.

Funds were then provided to employ a statistician and rigorous experimental design became an essential component of every sub-experiment. Formal experimental protocol development was introduced. Approximately 30% of all grains used in each sub-experiment were grains used in other experiments to act as 'connectivity' across sub-experiments. Partial replication of grain diet pelleting, individual animal recognition, spatial and temporal

variability were considered as potentially important factors in all future experimental designs.

## Value of experimental design

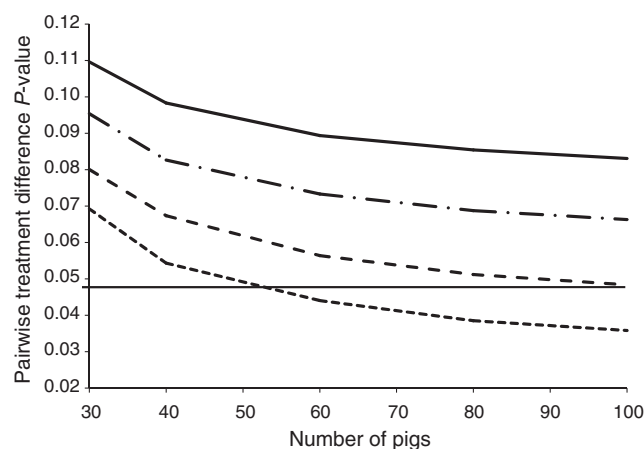
Two examples of the value of statistical design are presented. The first shows the importance for improving the accuracy of the estimate of digestible energy content of cereal grains for pigs of measuring and taking account of factors that may influence the variance of the primary measurement. Pelleting of the diets takes place over several days and although the procedure is recommended to be constant, the pellet press operator sometimes varies settings to enhance pellet quality. Thus, ~30% of the treatments were replicated at the pelleting stage to prevent pellet batch and grain source from being confounded when estimating the digestible energy content of the grains. The process of replicating only a subset of treatments is known as partial replication (see for example Smith *et al.* 2006).

The impact of experimental design on factors contributing to non-grain variation is shown in Table 1 for two sub-experiments. If factors that may influence the measurement of the digestible energy content of the cereal grains in the sub-experiments were not measured and accounted for during the statistical analyses, all of the variation in measurement would be allocated to residual. By applying a rigorous design and analysis, the residual error variance as a percentage of non-grain variation was reduced from 100% to 32% in sub-experiment 1 and to 20% in sub-experiment 2.

The consequences of not applying a rigorous experimental design are illustrated for sub-experiment 2, where only 20% of the non-grain variance was residual. If the design did not account for non-grain factors and we assume that the variance associated with these factors becomes part of residual error variance, then the residual error variance for the estimated digestible energy content of the grains is 0.1226 MJ/kg. However, when the above non-grain factors were considered in the statistical analysis, the residual error variance was reduced to 0.0252 MJ/kg. This reduction in residual error variance changed the error in the estimate of the digestible energy content of grains in the sub-experiment from  $\pm 0.35$  to  $\pm 0.16$  MJ/kg. Such a change more than doubled the potential accuracy of a NIR calibration for predicting the digestible energy content of cereal grains for pigs. In addition, by accounting for factors contributing to non-grain

**Table 1. Factors contributing to non-grain variation in two sub-experiments to determine the digestible energy content of cereal grains with partial diet replication in the pelleting phase of the experiment and an experimental design accounting for spatial (pig number, pellet batch) and temporal (pelleting day, feeding run, feeding period) factors**

Design category	Contribution to non-grain variation (%)	
	Sub-experiment 1	Sub-experiment 2
Pellet batch	48	39
Pelleting day	0	27
Pig number	17	6
Feeding run	0	0
Feeding period	5	8
Residual	30	20
Total non-grain variance	100	100



**Fig. 1.** *P*-values associated with detecting a pairwise difference in grain digestible energy content of 0.33 MJ/kg for different percentages of diet replication at the pelleting stage of the experiment and total number of pigs in the experiment. Solid line – 20% replication; dashed and dotted line – 50% replication; large dashed line – 80% replication; small dashed line – 100% replication.

variability and using a standard formula found in experimental design textbooks (Kuehl 2000), 5-times fewer replicates (animals) are needed to detect a 0.25 MJ/kg difference between two grain samples at the  $P < 0.05$  level of statistical significance compared with not accounting for these non-grain factors.

The second example illustrates the importance of statisticians in helping research managers make objective decisions about resource allocation in animal experiments. Optimising resource use and number of animals to obtain statistically significant differences between treatments has implications for animal welfare and for efficient use of research funds. Results from Table 1, sub-experiment 1, where pellet batch accounted for 48% of the non-grain variation, are used for the illustration. For this example, it was assumed there are 10 treatments (grains) and that the pellet processing facility could complete five batches of pellets a day. Fig. 1 shows the effect on *P*-values associated with detecting a pairwise difference in grain digestible energy content of 0.33 MJ/kg for different percentages of grain (diet) replication at the pelleting stage and total number of pigs in the experiment. The example shows that at least 80% of the diets must be

replicated to detect a pairwise grain difference between the grain samples at the 5% level of statistical significance. With 80% of diets being replicated at the pelleting stage, ~90 pigs in total (9 per treatment) are required to detect a significant difference in energy content between the grains. However, if all grain diets are replicated at the pelleting stage, only 50 pigs (5 per treatment) are required to detect the difference between the grains at the same level of statistical significance. The optimal combination of diet replication at the pelleting stage of the experiment and total number of pigs in the experiment can be determined from such an analysis based on the various costs of resources and availability of facilities.

## Conclusions

Outcomes from animal research and returns on invested research funds are likely to be enhanced considerably when greater attention is given to ensuring steps of the scientific method are rigorously followed and if research funding explicitly includes resources to enable statisticians to participate as integral and equal partners in any research project. Achieving these objectives is the joint responsibility of the scientists, research organisations and fund providers.

## Acknowledgement

The results analysed were from a project part funded by the Pork CRC.

## References

- Baldwin RL (1995) 'Modelling ruminant digestion and metabolism.' (Chapman & Hall: London)
- Black JL (2008) Premium Grains for Livestock Program: Component 1 – Coordination. Final Report. Grains R&D Corporation, Canberra, Australia. Available at <http://ses.library.usyd.edu.au/handle/2123/1912> [Verified 5 November 2015]
- Fisher RA (1938) Presidential address to the First Indian Statistical Congress. *Sankhya* 4, 14–17.
- Kuehl RO (2000) 'Design of experiments: statistical principles of research design and analysis.' 2nd edn. (Duxbury/Thomson Learning Press: Pacific Grove, CA)
- Levine TR (2013) A defence of publishing nonsignificant (ns) results. *Communication Research Reports* 30, 270–274. doi:10.1080/08824096.2013.806261
- Smith AB, Lim P, Cullis BR (2006) The design and analysis of multi-phase plant breeding experiments. *The Journal of Agricultural Science* 144, 393–409. doi:10.1017/S0021859606006319