Further aspects of the use of individual animals as replicates in statistical analysis

The issues raised in recent letters (Phillips, 1998; Rook, 1999; Weary and Fraser, 1998; Jason and Elston, 2000) discussing the use of individual animals as replicates in statistical analyses are of major significance for the future conduct of research with groups of animals, mainly farm animals. Several writers have indicated that they believe the relevance of the discussion is not just to animal behaviour responses, but because of the links between behavioural responses and animal production, reproduction and nutrition, to the whole range of animal production research. If we accept that animals should never be used as replicates, we cast doubt on the majority of farm animal research conducted in the twentieth century. It seems unlikely that such a coherent picture of animal responses to management, nutritional, reproductive and other treatments would have evolved in the twentieth century if the basic premise for statistical comparisons was wrong.

Weary and Fraser (1998) argue that the critical issue is whether treatments are applied to individuals or to groups of animals. However, if treatments are applied to groups, they are de facto also being applied to individuals. There is just the same chance that there will be an interaction between animals, when individuals are subjected to treatments, for example a different milking procedure for cows, as when a group of animals is subjected to a treatment, for example changing the cows’ ration. A different milking procedure such as automatic milking could influence the behaviour of cows at times other than during milking and in relation to conspecifics. A ration change may alter the behaviour of individual cows in a different manner, but again potentially disturbing the social structure of the herd. The key issue is not whether the treatment is applied to herds or individuals but whether an individual’s response is influenced by other cows and the extent of the interaction. If it is, then the treatment effect on that animal is not independent and the error term in the statistical analysis is not random.

Weary and Fraser (1998) broaden the discussion to criticise the use of genetically related members of the same herd to derive results that are applied to a much wider genotype. Rather than being too dogmatic in limiting the use of related animals as replicates, researchers must be aware of the limitations of the applicability of their results — and most are. It would be folly to suggest that responses of Holstein–Friesian cattle to temperature stress are applicable to Zebu cattle, for example. However, a cautious researcher might imagine that behavioural responses of Holstein–Friesian cattle to changes in light intensity will be at least qualitatively the same as the responses of Zebu cattle. Others may wish to take representatives of a number of species to derive...
Letter to the Editor

general physiological principles, for example the work of Heffner and Heffner (1992) on auditory perception in mammals. Weary and Fraser (1998) also fire a passing shot at the use of replication in time, which they also derogatorily refer to as pseudoreplication. However, this too has its place on occasions to confirm that a response within an individual or a group of animals is repeatable over time. As long as the researcher understands that the error term relates to variation over time, and not between individuals, it is an acceptable analysis if it is recognised that the results are only relevant to the individual or group of individuals used for the study. Replication over time can also be a legitimate component of animal research if, for example, the capacity of a building is limited, requiring several batches of animals to be subjected to the same treatments at different times. Providing other variables are fully controlled, which will be most applicable if the building is a controlled environment, then the measurements of responses to the same treatment in different animals kept at different times is legitimate (see for example, Prayitno et al., 1997). Weary and Fraser’s concern no doubt arises because some researchers have used replication over time because no measurements of variation between individuals were available.

Rook (1999) argues against my conclusion that for some behaviours individual animals should not be used as replicates on the grounds that “it is a recipe for confusion … and impractical since most studies that are likely to contribute significantly to our knowledge will require the measurement of behaviours that must be regarded as liable to social interaction until proved otherwise.” In other words, we assume that behaviours are socially facilitated until proven otherwise. This approach is impractical because behaviours are influenced to different degrees by social interaction. Some may occasionally involve slight interactions with other animals that may have minor implications for the independence of the individual animals. Others involve positive interactions, e.g. inter-individual aggression, which probably precludes the use of individual animals for statistical analysis. Negative interactions, as for example where an aggressive animal renders other animals less aggressive by virtue of their dominant behaviour, could reduce the chance of analysis of variance detecting differences between treatments — the conservative test described by Iason and Elston (2000).

A selective approach to the use of individuals as replicates is most practical and logical. If a researcher wants to report an experiment investigating largely involuntary behaviours such as blinking, urination or even milk letdown in dairy cows, it would be time-consuming and unnecessary to prove that they are not socially facilitated, before an experiment involving animals as replicates can be conducted. Conversely, an experiment investigating yawning in humans would clearly involve social interaction and could not use individuals as replicates. Therefore, we already know that some behaviours involve social interaction between animals and others do not, and this is also true of production, reproduction and physiological parameters. The researcher, and those refereeing the research reports prior to publication, should evaluate the likelihood of the results having been influenced by social interactions. All research reports should then be evaluated on the quality of the finding in relation to the robustness of the research methodology, acknowledging that no research technique is perfect. It is easy to forget that even the calculation of probability values acknowledges the possibility that an effect has occurred by chance. Adopting a rule as to which probabilities are acceptable, as certain journals
do, dissuades scientists from conducting exploratory science and encourages them to prove things that are expected and to many, obvious. The importance of a finding must be evaluated against the risk that the result occurred by chance. For example, a farmer might think that a research discovery of a 90% probability that a disease organism will mutate in response to climate change and kill all his cattle is of much greater importance than a discovery that there is a 95% probability that the same organism will mutate but only result in the death of one tenth of his animals. If the former observation is repeated, and by necessity reported, several times the greater significance of the finding, in comparison with the latter observation, is obvious.

What can we conclude in relation to the use of animals as replicates? Should we, as Rook (1999) suggests, still publish the results, as means or medians, if appropriate, and standard errors if the values are normally distributed, but exclude the results of any statistical test. This is tantamount to saying that the authors or the journal editors are not sure whether an observed effect is valid, so only the observation is reported with no indication of its significance. Readers will rapidly conduct their own assessment of the significance of the observation from the difference between treatment means and the standard errors. A better option is for researchers and journal editors to individually evaluate whether social interaction has influenced the results of each experiment.

In clear cases of social interaction preventing the use of individual animals as replicates, there are alternative approaches. The obvious one is to use groups of animals as replicates, as is often already done where animals are kept in small groups, e.g. battery hens. However, the conditions in which each group is kept must be identical or it could influence the results — a potentially serious problem with groups of large animals such as dairy cows. Gibb et al. (1999) have described such an approach for grazing cattle, where a herd of cattle is divided into groups of three animals, which then form the replicates for statistical analysis of treatment effects, in this case sward conditions. However, it is difficult to be certain that a group of three animals responds in the same way to treatment as a larger group, especially as the cohesive relationships between animals in the large herd will be broken when the groups of three are formed. Groups of four cows are more vigilant during grazing and less aggressive than groups of eight or 16 cows (Rind and Phillips, 1999). Again, a value judgement must be made for each piece of research, as to whether the small group size is likely to have influenced the response to treatment. Another possible approach is to identify cows within a herd that do not interact, and can therefore be considered as independent. This requires observations to be conducted before the experiment starts and the researchers cannot be sure that all possible communication between herd members has been monitored. Nevertheless, if one accepts that no approach is going to be perfect, this offers perhaps the most promise, especially as it retains the animals in a normal herd structure.

This debate will continue and almost certainly no definite solution will be found, but what matters is that researchers think about the constraints and possible inaccuracies in their experimental design. Science will not be served by the introduction of rigid rules governing the use of statistics; rather the significance of the results must be interpreted in the light of their potential impact and reliability.
References


C.J.C. Phillips
Animal Science Division,
Department of Clinical Veterinary Medicine,
University of Cambridge,
Cambridge CB3 0ES, UK
E-mail address: cjcp2@cam.ac.uk