

ANCCART News Facts Sheet

The importance of non-statistical design in refining animal experiments

David B Morton

Department of Biomedical Science and Biomedical Ethics
Medical School, University of Birmingham, UK

Introduction

Refinement, one of the Three Rs, is perhaps the area where most progress can be made to reduce any animal suffering incurred in animal experimentation in the short term. Another 'R', Reduction, will reduce the number of animals used in research through good statistical experimental methods and design. This paper shows how non-statistical strategies can contribute markedly to minimising pain and suffering in an experiment. The third 'R', the Replacement of animals in experiments, however, is a longer term aim which requires considerable investment, and input of basic animal biology data, before absolute animal alternatives can be made available to provide reliable scientific data. In the past 20 years or so, Replacement has developed as a concept more in relation to toxicity testing and finding ways in which the necessary harms done to animals (e.g., the necessity to produce clinical signs of toxicity in order to determine the toxicity of a chemical) can be eliminated. In fact, less than 10% of animal experiments involve toxicity testing (UK Home Office Statistics, 1996). The use of replacement alternatives in the other 90% of experiments is less likely as whole animals are needed. This is because the purpose of the experiment is not simply to determine toxicity but to determine function of part of the body (eg. the immune system, the reproduction cycle) or to test the effectiveness of new medicines (eg. antibiotics, anti-hypertensives, anxiolytics, analgesics, anesthetics) and so require whole animal body systems to be fully functional. Nevertheless, even in these situations, *in vitro* methodology can provide important information for the investigator. It may seek to answer questions such as whether a chemical interacts with cell receptors, or at what level is it toxic, or does it fail to interact?

Refinement

Refinement not only attempts to reduce negative states in animals, but promotes positive mental and physical states. It can be defined as follows:

Those methods which avoid, alleviate or minimise the potential pain, distress or other adverse effects suffered by the animals involved, or which enhance animal wellbeing.

The scientific paradigm is normally to vary one parameter at a time in some form of standard experimental protocol. Thus identifiable variables are limited as much as possible in any one experiment, so that only the specific variable of interest is examined. This objective will be compromised not only by poor experimental design but also through inadvertent side effects which can occur as a result of compromised animal wellbeing. Unintended animal suffering deriving from physical damage to an animal, or physiological or psychological perturbances from normal, can materially affect the scientific data collected and so should be

avoided as far as possible. Avoiding such confounding effects is more than simply the provision of anesthetics and analgesics when animals may be in pain, or alleviating distress and fear. While these are very important, there are additional strategies which can be employed to reduce the level of animal suffering, notably the statistical and non-statistical design of experiments. Traditionally, research into the refinement of experiments has not been rigorously investigated as it is not seen to be part of the main line of research. Consequently, important questions such as how much pain should be caused to demonstrate effective analgesia are often ignored. If scientists wish to claim they practise humane science then they have to pay as much attention to avoiding unnecessary pain and distress to their animals as they do to their scientific objectives.

Humane science will influence the welfare of the animals as well as the scientific quality of the work, and deserves detailed consideration. Moreover, attention to refinement will have a beneficial effect on the costs of the research, because both experimental design and the welfare of the animals affect the variance in the scientific results. Poorly designed experiments will be likely to use more animals than necessary to achieve the scientific objective, which in turn will lead to increased costs for upkeep, administration, record keeping and treatment. As important scientific discoveries often have to be validated by replication of the work in another laboratory, this too will add to the overall costs if the work has been poorly carried out, especially if not accurately and fully reported. It is important that scientific protocols, including the precise experimental design, are recorded in detail in peer reviewed journals (Morton, 1992). Factors leading to good animal welfare will usually also lead to reliable, accurate and economic science.

Literature review and choice of model

There are three primary points to establish before embarking on any experiment. Has the work been done before; is there an alternative way of achieving the scientific objective without the use of sentient animals (eg. using invertebrates, or in vitro techniques); and is the scientific question worth answering, in the light of the pain and suffering that may be caused to the animals during both their husbandry and the experimental procedures.

It is also important to review the literature before starting any work, focusing not only on whether the proposed approach to the scientific objective is appropriate, but carrying out a critical analysis on whether the proposed model really will achieve the stated objectives. If there is a choice of models then the advantages and limitations of each should be considered. For example, is the model simply reproducing the clinical signs of a disease, or reflecting the actual cause of the disease, such as the infective agent or the genetic defect (it is in this area that transgenic animals should be improving the validity of certain animal models of human diseases)? Has the model been validated in any way, if so how, and is it scientifically acceptable?

The choice of species can be critical when the results are to be extrapolated to humans. Those animal species with similar genetic predispositions, or metabolic pathways, or that show similar pathology to humans are likely to be better models. When the purpose of the work is specifically aimed at a non-human species (eg. veterinary, zoological or wildlife research) then other considerations come in, such as the stresses involved in capture or confinement. It is also important to remember that there are strain or breed differences within a species which can be critical (Claassen, 1994; Hendriksen and van der Gunn, 1995). Both species and strain should be comparable for scientific validity.

Other questions include whether adequate facilities are available locally to create the model. Is the necessary expertise available to recreate it accurately and precisely; and are the staff able to care for the animal afterwards and during its lifetime (Smith and Boyd, 1991)?

In some countries special justification is required for the use of some species, eg. in the UK the use of dogs, cats, horses and primates. This is not related to the biological needs of these animals, or to their ability to feel pain more than other species, rather it is an attempt to reflect the special public concern over the use of some species compared with others. Those species that are often eaten, regarded as pests, or are at least distasteful in some way or another, even though they may also be kept as pets (eg. mice, rats, gerbils, hamsters, guinea-pigs, rabbits) cause least public concern. Surely, the rational choice of species is the one that is the least sentient consistent with sound science.

The use of anaesthetics and analgesics

Scientific procedures should be performed under anesthesia (ie. general, regional, local or with analgesia) unless either the pain or discomfort of the procedure is less than that caused by the anaesthetic; or its use would be incompatible with the scientific objectives. The latter has to be shown and not assumed and a scientific approach is the best way to do this. Often assumptions are made to the detriment of animal wellbeing, and possibly also to the quality of the science, because of the animals' responses to stressors; but by not researching this, it is never revealed. In addition the type of anaesthesia should be suitable for both the species and the duration of the procedure and the researchers should be competent in its administration. If no anaesthesia is to be used it is important that pain and fear should be kept to a minimum through the use of analgesics and tranquillisers.

The importance of competent and skilled investigators

Competent and skillful researchers are vital for the production of good science. Poorly carried out techniques can cause considerable, and avoidable, suffering and it is in the interests of both science and animals that all scientific procedures are carried out in the most humane manner. A new technique that demands new skills should be carefully learned by an investigator before embarking on an experiment in which animals recover and data are recorded, even if more animals have to be used. Becoming competent in a technique involves watching those who are competent and learning the 'tricks of the trade'. Depending on the technique to be learned, it may be appropriate to dissect dead animals to learn the anatomy and to understand the inherent dangers of inaccuracy. The equipment to be used must be suitable to the species, strain, sex and size or age of the animals. For a surgical technique, the next stage may be to practise on recently killed animals so that the tissues are warm as this can markedly alter their texture and hence the value of the learning experience for the operator. Finally, a surgical technique can then be applied to anaesthetised animals which may be allowed to recover if it has gone well, but if not, those animals should be killed. The importance of learning with a competent, skilled and up-to-date scientist or technician cannot be overemphasised. Moreover, even simple techniques on conscious animals like handling, oral dosing, parenteral administration of substances and removal of blood, if carried out badly, can significantly elevate the level of animal pain and distress. This in turn will cause further problems next time as the animal anticipates what is about to happen based on its previous experiences.

Statistical advice

It is well worth obtaining statistical advice before an experiment is started. There are many standard texts and commentaries (eg. Mead, 1988; Erb, 1990; Mann *et al.*, 1991; van Zutphen *et al.*, 1993; Festing, 1994; Erb, 1996; Chamove, 1996; Khamis, 1997) covering the various approaches. They include concepts such as avoiding using too few animals during an

experiment as opposed to too many; what data are to be collected and how the data should be analysed; the value of pilot studies and prior information of expected variance; the calculation of the number of groups, the number of animals in each group and increasing the power of an experiment through blocking; ongoing analysis during an experiment; as well as statistical instruments and tests.

The limitations of scientific evidence, extrapolation and animal numbers

The purpose of an experiment has to be clearly defined, including the primary application of the results. For instance, in safety testing for a new chemical entity, the necessity to obtain acute toxicity data to a probability of <0.001 is pointless when an extrapolation is to be made to another species, normally the human. Moreover, if a fudge factor, such as a further safety margin of 100x the toxic dose found in animals, is going to be used to calculate the hazard involved in any exposure for humans, say for the transport of that substance, then the futility of such precise experimentation is highlighted even further. The question to be asked is, What is going to be done with the results? When there is a clear difference in consequent actions based on the difference in probability estimates, and the preferred action can be justified, the required animal use is acceptable. But if there is to be no difference in action between different statistical precisions then the lowest acceptable precision which will lead to the use of the fewest animals should be used.

Another instance is when it has been decided that even if one animal in a group fails a test, then a specific action will take place. For example, if a substance is going to be labelled as toxic on the basis of one animal in a group of six reacting adversely, and if the first animal reacts adversely, then there is no point in continuing with the experiment and giving that substance to the remaining five animals. It would have provided valid scientific information, of course, but the practical applied outcome has already been decided and so there is no point in continuing the study and causing pain and suffering to the remaining animals.

Protocols evaluating new treatments, eg. the rodent protection test (RPT) help determine, for example, effectiveness and therapeutic levels of antibiotics for use in humans. In the RPT, the effectiveness of a novel compound to prevent or treat infection is investigated over a range of doses using standard micro-organisms in an animal model (mouse or rat). If, compared with others similar drugs, the compound appears to have advantages (eg. showing a better bacterial kill rate or less toxicity) then the information gained from the animal studies will be taken into account when Phase one (healthy human volunteers) and Phase two (human patients who can expect to benefit therapeutically and in whom effective dose levels will be determined) studies are undertaken. There is little need for high precision in the animal model species as opposed to the target species and yet the approach of accepting a lower statistical probability ($p < 0.05$ cf. $p < 0.01$) may mean the difference between 5 and 50 animals in a group. This is especially important in the area of infection where considerable suffering may occur in control animals which are not protected from the disease.

Pilot studies, dose sighting, and control groups

Pilot studies can be useful to give some idea of the variance, the likely adverse effects that may be encountered, and any practical local problems that are likely to be encountered during the proper study. They help to determine the appropriate number of animals that should be used in any one group, as well as the avoidance and alleviation of any adverse effects. The use of pilot studies need not be a waste of animals, as the results can be used for the main study, unless there are good reasons for not so doing.

Historical data in this regard are also very useful, particularly when considering both positive (eg. a standard challenge of a known substance) and negative (eg. giving vehicle alone) controls. This use of background data can influence the number of animals needed for a given study. In some instances it may not even be necessary to have a control group as the control 'fact' is so well known and accepted eg. total pancreatectomy leads to diabetes mellitus with a consequential rise on blood sugar and death in less than 10 days. However, it may be advisable to carry out pilot studies when using a new model in one's own animal facility as differences in strain of animal, diet, staff, bedding, cleaning materials, husbandry and environment have all been shown to affect animals' responses in various ways (see Claassen, 1994; Morton, 1995; Wadham, 1996). This was exemplified by a multicentre trial on the Fixed Dose Procedure carried out by van den Heuvel and co-workers (1990). They found that while the LD50% of various chemicals were similarly ranked between laboratories, the actual LD50% dose varied considerably.

There are several methods for determining the relevant dose(s) for animals *in your facilities* in pilot studies but the up and down method deserves special mention (Bruce, 1985). In this approach, a single animal is given a dose (eg. from the literature) and its response noted. Assuming that the first dose does not produce the anticipated response, a second animal is given a higher or lower dose. This is then repeated on individual animals at varying dose intervals until a suitable dose rate is found. This approach is helpful for streptozotocin dosing to induce diabetes in various strains of rats, as well as for other chemicals given to induce various disease states.

Advice from other scientists, veterinarians and caretakers on adverse effects

The first time a novel scientific procedure is carried out will require a prediction of the type and incidence of any adverse effects on the animals, but thereafter the experience gained will be helpful in future work using that or similar protocols. While a thorough review of the literature is essential, sadly the adverse effects on the animals encountered are rarely published in a helpful manner for those using the model (see Morton, 1992). Furthermore, the 'tricks of the trade' so often vital to be able to repeat the work are omitted, let alone relevant details relating to the animal side of the work, e.g., failed approaches, details of animals used as well as the husbandry system. The use of score sheets (Morton, 1994, 1995, 1998b; Townsend and Morton, 1994; Morton and Townsend, 1995) help all those involved in the research program to recognise the cardinal clinical signs that animals show during a specific scientific procedure, and can signal subsequent actions, notifications and treatments to be taken when certain clinical signs are observed (eg. humane end points). Animal carers, stockpersons and veterinarians who are familiar with the animals in the experiment are excellent guides in determining any adverse effects. They may not always be able to explain why an animal is 'wrong' or 'not right' but good ones are rarely mistaken! The score sheets start to explain their insight and simple tests like giving an analgesic and observing if animals change their behaviour such as by moving around more, eating and drinking, all add empirical evidence to suggest those animals were affected by that protocol and suffering in some way (see below for more details).

In vitro to in vivo

Research is sometimes aimed at investigating the interactions of substances, such as new drugs, in the whole body. If the *in vitro* work will yield information that shows the effectiveness of the substance, or viability of the cells/organisms has been compromised in some way, it may not be worth proceeding with the *in vivo* work. There has to be confidence that the *in vitro* work is able to be extrapolated to the *in vivo* situation, and if so the *in vitro* work should precede the *in vivo* experiments. For example, genetic modification of viruses may reduce their virulence so that they

are no longer able to infect cells *in vitro* and, based on past experience, it is unlikely they will infect cells *in vivo*. Similarly, radiolabelling of bacterial toxins may remove their toxicity, which can be measured by a change in binding characteristics to monoclonal antibodies, or to cell surfaces *in vitro*, or to a change in toxicity to less sentient animal forms such as invertebrates. Again, based on experience, it may be unlikely the modified toxin will be effective *in vivo* and so these experiments need not be done. In cases of doubt, however, small pilot projects involving only two or three animals can be used (depending on the precision needed) to confirm such findings and so validate a predicted negative, especially when there is less likelihood of causing animals pain and suffering.

When a positive harmful effect needs to be confirmed *in vivo* then only one or two animals may be needed as *any toxicity* will be likely to provide sufficient evidence. For example, it may render a product such as a shampoo causing corneal ulceration or even corneal rupture unmarketable. Furthermore, in such cases where an adverse effect is predicted and needs to be confirmed, it may be possible to prevent the animal feeling any pain by carrying out the work under general or local anaesthesia.

Specific tests may be carried out to determine if an experiment is likely to work *in vivo*, or whether it is unlikely to succeed for some reason. If a chemical is cytopathic i.e., toxic to cells *in vitro* at concentrations that were predicted to be useful *in vivo*, then the work should not proceed. This is the basis of the so-called pre-screening tests used in the early stages of toxicity testing of new chemicals and medicines. An example of this might be a chemical in solution which at the desired concentration has a pH which is far from being biocompatible and will obviously cause tissue damage if given parenterally. It may then be given by some other route e.g., enterally, or at a lower concentration or in a different formulation. In all these cases consideration is given to carrying out *in vitro* work before whole animal work in order to avoid, or alleviate better, any adverse effects in animals.

Within in vivo work

Even within *in vivo* work it may be possible to stage work so that serious adverse effects are avoided. For example, if one is looking at the survival of animals given a bowel transplant, the animals are likely to die painfully from a peritonitis. Whilst one could wait for clinical signs of peritonitis to become apparent, it is also possible to avoid that altogether by trying to pick up signs of very early rejection, rather than bowel perforation (caused by rejection leading to death of intestinal wall cells and disintegration of the bowel). The early signs of transplant failure may be thrombosis of the vessels leading to the graft within a few hours of the end of surgery. By keeping the animal under a terminal anaesthetic for say 12 hr or so, one can monitor blood flow and the condition of the transplant (colour and contractility) and if found to be deleteriously affected, the animal can be killed before it recovers from the anaesthetic. If successful after 12 hr or so, the animal can either be permitted to recover, or I would recommend that the next animal be permitted to go for 24 hr before being killed. In such a way animal suffering can be reduced considerably, and it may be possible to avoid serious adverse effects altogether.

Another approach is always to do key pilot experiments before any controls or further experiments. Experiments should be designed to provide crucial information on whether it is worth progressing down certain experimental routes. For example, if the experiment is to determine the effect of a substance or specific cell or tissue on a physiological response, then a pilot study showing first that that response actually occurs should be undertaken. Only then should the main study and necessary controls, such as using saline or sham operations, be carried out.

Mild to high severity

Several examples can be instanced where the harms done to an animal can first be shown to be

effective or ineffective at a low level before going to higher levels of animal pain and suffering. Traditionally this aspect of refinement has not been rigorously investigated as it is not seen to be part of the main research, and the 'standard/traditional' time/level normally used. In some psychological experiments animals may be starved for long periods in order to make them 'work' in some way or another. If the animal can be motivated to work by reward rather than by starvation or even aversion stimuli, then such a strategic progression of harm could reduce the amount of animal suffering without losing the scientific objective. If starvation is used to motivate the animal then how long should it last - 48hr, 24hr, 18hr, 12hr or 6hr - and does it depend on the age of the animal, the species or strain? The adverse physiological effects of such periods of starvation (and even dehydration) on the animals have also to be considered and whether this will affect their performance and hence the science. If aversion has to be used, then the same principle should apply: use the lowest level of aversion (e.g., electric current, sound level, learned helplessness) necessary to achieve the scientific goal. Testing of novel analgesics, or determining neurocircuitry or transmitters may be possible through causing low levels of painful stimuli rather than high ones. An analgesic ineffective at low levels of pain is unlikely to be effective at higher levels of pain. The neurobiology may change quantitatively but not qualitatively at the higher levels of nociception. In any event, the purpose of the experiment has to be explicit, and justification made for the higher pain levels. Finally, the purpose of dose sighting studies mentioned earlier and practised by the toxicologists in safety testing are not routine in other areas of science. A scientist may start a dose response experiment at a range of dose levels only to find all the animals are unaffected, or worse still all the animals have died. Pilot studies are always worth doing in these sorts of cases.

Other examples include only affecting the use of one of a paired organ, that is important for an animal's quality of life, rather than two e.g. legs in fracture or arthritis or musculoskeletal research; sense organs such as eyes and ears; transplanting a third kidney into an animal in the groin where it can easily be monitored rather than removing an animal's own kidneys and replacing it with a donor kidney (this could also be an example of a staged approach if true dependence needed to be established). A similar approach can be adopted in the study of skin transplantation or burns research where a minimum area should be transplanted or damaged or to neural studies where nerve section or denervation is required: rather than incapacitating the whole leg through high nerve section, a low nerve section which will affect just one muscle in the distal part of the limb so that compensation is possible through the use of other muscles, should be carried out. In all these examples if the purpose is to study quantitative aspects then this can be staged so that small harms are first seen to be 'effective' in whatever scientific sense is appropriate, and then the insult can be scaled up providing the scientific justification is ethically acceptable.

Cancer research

A set of excellent guidelines exist for the refinement of experiments involving cancer (UKCCCR, 1997). The recent increase in cancer experimentation as a result of gene therapy work makes this an important area for refinement. In essence, the growth rate of tumors and their size should be kept to the minimum necessary to demonstrate effectiveness or ineffectiveness of novel therapeutic agents or regimes. Death can hardly ever be justified as an end point (Mellor and Morton, 1997). Tumors should be placed whenever possible in sites where they are easy to palpate or monitor. Animals should be monitored at least daily, and more frequently if necessary e.g. at times when the tumor ulceration is likely to occur or the animal may be found dead. Other parameters indicating animal wellbeing should be sought as well as tumor size, particularly if the tumor is deep or invasive or internal and so difficult to detect. These may include the behavior of animals, their coat quality, posture and appearance, especially their behavior at night, a period when they should be maximally active.

Humane end points

It is often possible given sufficient familiarity and experience with an animal model or specific scientific procedure that the adverse effects on animals will be predictable. This will not only include the clinical signs an animal might show but also the duration of those signs, the progression to other signs, even death, the timing of such signs after a particular procedure has been carried out, and also the proportion of animals that may be affected. Actions can then be taken depending on the scientific purpose and also on humane considerations. Certain levels of harm may be integral to the procedure or unavoidable to achieve the scientific objective. However, if the animal, because of the deviation from normality either physiologically or psychologically, loses its scientific utility, then it should be humanely killed as keeping it alive will yield no valid information. Similarly, if the scientific objective has been achieved, then there is no useful purpose in keeping the animal in pain or distress. If, however, it may yield useful information then it may be possible to alleviate the animal's condition and so reduce the level of suffering; or the level of suffering may have become incompatible with the expected degree of scientific benefit and on an ethical harm-benefit analysis, the animal should be killed. Alternatively, the animal may still have utility in a scientific sense but there was no need to have let the animal progress that far as earlier clinical signs could have predicted what was to come. This idea of using early clinical signs to predict later ones requires validation studies where it is shown that animals will normally progress in that way and that such an end point can be relied upon. This may be important for safety studies in the testing of medicines such as vaccine potency testing, rodent protection test, virulence assessments, or batch testing of natural or synthetic products. Cussler (1997) has recently presented some data on a lethal rabies vaccine potency test where he found that vaccinated mice given challenge doses of virus went through a series of predictable clinical signs. He showed that animals showing slow circular movements invariably progressed to death - the traditional end point for the test. Death, as so often is the case in many experiments, is not related to the experimental variable under study, but is due to indirect effects such as dehydration caused by animals not being able to drink and increased packed cell volume leading to heart failure. Another cause of death in mice can be low body temperature due to inadequate food intake. Sometimes death due to these indirect causes can take several days and Cussler in his work, and Sothillet *al.*, (1992) showed that animal suffering could be reduced by several hours to three to four days through implementing validated pre-lethal endpoints.

In the UK, under the *Animals (Scientific Procedures) Act 1986*, each scientific procedure has a severity limit which should not be exceeded, and indeed it is a criminal offence to do so. The question is how that limit is recognised in practical terms. In theory there are four recognised severity limits: mild, moderate, substantial and severe, but severe pain or severe distress is not permitted under any circumstance. In a sense, what the bands are called is irrelevant, what matters is how they are interpreted. In order to interpret them accurately and reproducibly, careful observation of animals is required, and score sheets documenting clinical signs with time are invaluable (Morton and Griffiths, 1985; Morton, 1985; 1986; 1990; 1994; 1995b; 1997a; 1998a; 1998b; Morton and Townsend, 1995). Severity limits can be interpreted as the degree of deviation from normality coupled with other indicators of health and quality of life. A body weight loss of up to 10, 20 or 25%, or greater than 25%, compared with controls may reflect the mild, moderate, substantial and severe limits. But animals with tumours or ascites may increase in body weight but lose body condition (ie. muscle mass) and be experiencing extreme suffering. Alternatively, animals with a body weight loss of 25% and which are diabetic or which have exocrine pancreatic deficiency, may be very lively and have a good quality of life. It is important that a holistic approach be taken so that clear clinical signs can be used to determine humane end points in accordance with the scientific benefit and humane research.

Score sheets - an approach to the recognition and assessment of animal suffering

Many laboratories are approaching the difficult topic of the recognition and assessment of suffering using score sheets i.e. a list of cardinal clinical signs encountered in that particular scientific procedure. These are developed through experience and, by and large, are unique to the system of husbandry, to the specific experiment, as well as to the species, and even the breed or strain of animal, being used. It is not possible to make a general score sheet for all experiments and/or for all species. One only has to consider the different potential adverse effects of a skin transplant compared with a kidney or heart transplant to appreciate the different signs that might occur (eg. haemorrhage, rejection of skin compared with a dependent kidney or an accessory heterotopic heart in the abdomen). Lists of signs are developed by observing the first few animals undergoing a novel scientific procedure very closely and then the list is modified with experience until a set of cardinal signs that most animals will show during that experiment, and that are relevant to the assessment of suffering, is achieved.

These key clinical signs are set out against time in the score sheet (see Table 1). On the left hand side are listed clinical and behavioural signs and along the top the days and the time of the recorded observations. The method of scoring is that clinical signs can only be recorded as being present or absent indicated by a plus or a minus sign (or sometimes a +/- if the observer is unsure). The convention is that negative signs indicate normality or within the normal range, and positive signs indicate compromised animal wellbeing. In this way it is possible to visually scan a score sheet to gain an impression of animal wellbeing: the more pluses, the more an animal has deviated from normality with the inference that it is suffering more than it was earlier. Clinical treatments and other observations are also recorded. It is important to note that animals can be scored at any time and it would be certainly be more than once daily during critical periods when an animal's condition could predictably give rise to concern (e.g., in the immediate post-operative period; in a study on infection after the incubation period and at the time of bacteraemia or septicaemia).

Practically, it is important to develop a disciplined approach and strategy to the recognition of adverse effects in animals. At the beginning of an assessment the animal should be viewed from a distance, and its natural undisturbed behaviour and appearance are noted. Next, as the observer approaches the cage or pen the animal will inevitably start to take notice and interact with the observer and that interaction can be used to determine whether it is responding normally (an animal may become inquisitive or show signs of fear). Finally, a detailed clinical examination can be carried out by restraining the animal in some way and observing its appearance carefully and then making clinical measurements eg. of body weight and temperature, in addition to its behaviour as it may be more aggressive or fearful, or even vocalise.

At the bottom of the sheet there are guidance notes for the animal technicians about what they should provide in terms of husbandry and care for animals on that scientific procedure. There are also guidelines on how to score qualitative clinical signs such as diarrhoea and respiration, as well as criteria by which to judge humane end points. Finally, if an animal has to be killed there is guidance about what other actions should be taken, such as tissues to be retrieved and kept in formal saline to ensure that the maximum information is always obtained from any animal study.

While these sheets take time to fill in it is not difficult for an experienced person, such as an animal caretaker, to see if an animal is unwell, so the time taken can be reduced by simply scoring that the animal is normal by ticking the NAD box (Nothing Abnormal Detected). However, if an animal is not normal, it does take time to check it and to make judgements over

what actions to be taken, but is that not the price for practising humane science?

In order to promote good care and good continuity of care an animal technician can be responsible for liaising with scientists and other technical staff, and also to maintain and update the score sheets. The roles of the technician in charge are:

- to check that the appropriate licences and permits are in order and correspond with what the scientist actually intends to do to the animal(s);
- to check the score sheet is appropriate before the experiment begins;
- to know the purpose of the experiment and the scientific objectives, and to become familiar with the scientific procedures to be carried out on the animals and the clinical signs that may occur;
- to ensure all personnel (technicians, scientists) know how to use score sheets and can recognise the clinical signs and can interpret these signs clearly into humane end points;
- to check that technicians not familiar with that experiment, say doing a weekend or holiday shift, are informed about the animals;
- to liaise with researchers over the experiment eg. timing, numbers of animals, equipment, end points;
- to update the score sheets based on new signs or combinations of signs observed; and
- to report to the responsible persons any concerns over the animals or personnel involved.

Table 2 shows a completed score sheet. At a glance it can be seen that there are more pluses to the right than to the left. Several other points can be noted: first, along the top, that as the animal became unwell, so it was scored more frequently. During Day 0 (the day of the operation) it scored abnormal in one or two predictable signs as it was recovering from the anaesthetic and the surgery (low body temperature and hunched) and so the NAD box was ticked. The next day (21st January) basic observations were made of amount of food eaten, temperature and body weight, and again the NAD box checked. However, towards the end of that day, the coat became starey (ruffled), the body temperature rose, and the breathing became more rapid. By the next morning, there was a significant body weight loss (12%) which increased during the day to 18% - a strong indication that the animal had not eaten or drunk much, if anything, and that it probably had diarrhoea. In fact there were so many abnormal clinical signs that it was decided to kill the animal on humane grounds before the end of the experiment. The sudden appearance of diarrhoea and the concomitant rapid weight loss and dehydration, laboured breathing, posture, lack of a red light response, etc all confirmed that the animal was becoming severely physiologically compromised and was not going to yield valid results in relation to the scientific objective. Even more significantly, its temperature was now at 35.1°C - a very poor sign, and the extremities were blue (i.e., the colour of the feet and ears). In our experience, this animal would have died that night if not sooner.

This scheme of scoring clinical signs for the recognition and assessment of adverse effects on animals during scientific procedures has been shown to have several advantages and include the following.

- Closer observation of animals is now carried out by all staff at critical times in the experiment as the sheets have indicated those times that are critical for the animal, and

when the animals find their circumstances mostaversive.

- Subjective assessments of suffering by staff and scientists are avoided,thereby promoting more fruitful dialogue, as evidence-based opinion becomespossible based on the clinical signs. In a sense they empower the techniciansand help them illustrate to less experienced persons why an animal is 'notright'.
- Consistency of scoring is increased as the guidance is clear and thescoring options are limited.
- Single signs or combination of signs can be used to indicate overallseverity of the procedure, as well as alleviative therapies or scientificprocedures at set points in the experiment (e.g., blood sampling).
- The sheets help to determine the effectiveness of any therapy intendedto relieve adverse effects.
- The sheets help to determine which experimental models cause leastpain and distress, e.g., by comparing alternative animal models helpingin refining the scientific procedures.
- The sheets help to train those inexperienced in the assessment ofadverse effects.

This has led to better animal care as well as providing useful scientificinformation such as the recognition of neurological deficits, times of epilepsyor weight loss, as well as unexpected findings such as urinary retentionin a model of renal failure. Furthermore, by picking up signs of poor animalwellbeing early, we can implement humane end points sooner rather than later,which avoids animals being inadvertently lost from an experiment. In theUK, where severity limits are imposed on each scientific procedure, thesheet can be used to indicate when such limits have been breached, or areabout to be breached, or may have to be reviewed, by the precise observationof the presence of clinical signs. The score sheet system provides a visualaid, opens up discussion between interested parties, and helps focus attentionon to the animals' condition throughout the procedures. An analysis of thescore sheets can reveal patterns of recovery or deterioration and so givesa better picture of the effect of a procedure on the animals from startto finish. The sheet encourages all involved to observe the behaviour ofanimals, to recognise normal and abnormal behaviours, which will help indetermining animals' responses to various procedures, and this will helpthe development of refining experimental technique by highlighting the typeand timing of any adverse effects. The scoring system has proved to be especiallyuseful with new procedures, or when users are not always sure of what effectsa procedure will have. In my experience the literature rarely records adverseeffects on the animals, or how to avoid them or measure them and I believescientists have a moral obligation to do so (Morton, 1992).

We now look more closely at ways of improving our peri-operative care andin some experiments we have found that recovery is slower than it couldbe if different anaesthetics or analgesics, or intraoperative proceduressuch as maintaining body temperature or giving warm saline were used. Wenow try to operate earlier in the day so animals have maximum time underclose observation and they can be given more support such as fluid therapyor special diets (eg. jelly, fruit, vegetables) which has proven to saveanimals lives as well as improve the speed of recovery. Finally, obtainingprofessional advice by seeking the opinions of experienced laboratory animalveterinary surgeons and animal technicians and stockpersons can be veryhelpful in determining what clinical signs should be looked for.

There is a philosophical debate about whether it is better to cause more suffering to fewer animals or less suffering for many animals to achieve a scientific end. That situation is not common in practice but the UK law takes the view that the level of individual suffering is what matters and so harms should always be minimised. Higher levels of suffering have to be justified on scientific grounds and not on the saving of other animal lives.

Concluding remarks

Refinement may end with the death of the animal but there are other important situations in which animals can suffer, such as their husbandry, social groupings, breeding conditions, handling, restraint, transport, and euthanasia. It is important that these are addressed for reasons of good welfare and good science (see Gunn, 1993; Tuli, 1994; Wadham, 1996). Much research is needed into the Three Rs especially refinement, and grant awarding bodies should be aware of their responsibilities in this regard not only to review carefully the scientific work, they fund, but also to fund research into refinement. Scientists must also write up their work fully and promptly in order to avoid others repeating the work even after a literature search.

Some may argue that the intensive monitoring of animals described above adds to the cost of the research, and of course it does. But scientists have a moral duty to inflict as little suffering as possible and such refinement is the price we should be willing to pay as a compassionate society, and as humane scientists.

References and further reading

Acred, P., Hennessey, T.D., MacArthur-Clarke, J.A., Merrikin, D.J., Ryan, D.M., Smulders, H.C., Troke, P.F., Wilson, R.G. and Straughan, D.W. (1994) Guidelines for the welfare of animals in rodent protection tests. *Laboratory Animals* **28**: 1-8.

Amyx, H.L. (1987). Control of animal pain and distress in antibody production and infectious

disease studies. *Journal of the American Medical Veterinary Association* **191**:1287-1289.

Arena, P.C. and Richardson, K.C., (1990). The relief of pain in cold-blooded vertebrates. *ANZCCART News* **3**(1):1-4.

Bach, S., Morten, F. N. and Tjelliden, N. U. (1988). Phantom limb pain in amputees during the first 12 months following limb amputation, after preoperative lumbar epidural blockade. *Pain* **33**: 297-30.

Balls, M., Goldberg, A.M., Fentam, J.H., Broadhead, C.L., Burch, R., Festing, M.F.W., Frazier, J.M., Hendiksen, C.F.M., Jennings, M., van der Kamp, M.D.O., Morton, D.B., Rowan, A.N., Rusell, C., Russell, W.M.S., Spielmann, H., Stephens, M.L., Stokes, W., Straughan, D.W., Yager, J.D., Zurlo, J. and van Zutphen, L.F.M. (1995). The Three Rs: The Way Forward. The Report and Recommendation of ECVAM Workshop 11. *ATLA* **23**: 838-866.

Blom, H.J.M., Van Tintelen, G., Baumans, V., Van Den Broek, J. and Beynen, A.C. (1995). Development and application of a preference test system to evaluate housing conditions for laboratory rats. *Applied Animal Behaviour Science* **43**: 279-290.

Brain, P.F. (1992). Understanding the behaviours of feral species may facilitate design of optimal living conditions for common laboratory rodents. *Animal Technology* **43**:99.

Bruce, R.D. (1985). An up-and-down procedure for acute toxicity testing. *Fundamental & Applied Toxicology* **6**:151-157.

Chamove, A.S. (1996). Reducing Animal Numbers: Sequential Sampling. *AWIC* **7**:3-6

Claassen, V. (1994) Neglected Factors in Pharmacology and Neuroscience Research. Biopharmaceutics, Animal characteristics, Maintenance, Testing conditions. In: *Techniques in the Behavioural and Neural Sciences*. (Ed. J.P. Huston). Volume 12. Elsevier, Amsterdam.

Cussler, K. (1997). Impfstoffprüfung: Kann die Bestimmung der Lethalitätsrate bei Belastungsversuchen durch eine Ermittlung klinischer Endpunkte ersetzt werden? Tierlaboratorium 20. Paper presented in Berlin at the BgW meeting May 1997. (Vaccine quality control - Is it possible to replace lethality in challenge tests by the use of clinical end points?).

Erb, H.N. (1990) A statistical approach for calculating the minimum number of animals needed in research. *ILAR News* **32**:11-16.

Erb, H.N. (1996). A non-statistical approach for calculating the optimum number of animals needed in research. *Lab. Animal* **25**:45-49.

Festing, M.F.W. (1994). Reduction of animal use: experimental design and quality of experiments. *Laboratory Animals* **28**: 212-221.

Fitzgerald, M. (1994). Neurobiology of Foetal and Neonatal Pain. In: *Textbook of Pain*. Eds. Patrick Wall & Ronald Melzack London: Churchill Livingstone. 3rd Edition. pp. 153 - 163.

Gunn, D. (1994). *The evolution of welfare in the husbandry of laboratory rabbits*. PhD Thesis, University of Birmingham, UK.

- Gunn, D. and Morton, D.B. (1995). Twenty-four study of the behaviour of New Zealand white rabbits in cages. *Applied Animal Behaviour Science* **45**: 277-292.
- Hendriksen, C.F.M. and Koeter, H.W.B.M. (1991). *Replacement, reduction and refinement: present possibilities and future prospects*. Elsevier, Amsterdam. pp 281.
- Hendriksen, C.F.M. and van der Gunn, J. (1995). Animal Models and Alternatives in the Quality Control of vaccines: Are In Vitro Methods or In Vivo Methods the Scientific Equivalent of the Emperor's New Clothes? *ATLA* **23**:61-73.
- HMSO (1996). *Statistics of Scientific Procedures on Living Animals in Great Britain*. HMSO, London.
- Khamis, H.J. (1997) Statistics and the Issue of Animal Numbers in Research. *Contemporary Topics* **36**: 54-59.
- Mann, M.D., Crouse, D.A. and Prentice, E.D. (1991). Appropriate Animal Numbers in Biomedical Research in Light of Animal Welfare Considerations. *Laboratory Animal Science* **41**: 6-14.
- Mead, R (1988). *The Design of Experiments*. Cambridge University Press, Cambridge NY.
- Mellor, D.J. and Morton, D.B. (1997). Humane Endpoints in research and testing. Synopsis of the workshop. In: *Animal Alternatives, Welfare and Ethics*. (Eds. LFM van Zutphen and M. Balls). Elsevier Science BV, Amsterdam, The Netherlands. pp. 297-299.
- Morton, D.B. and Griffiths, P.H.M. (1985). "Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment" *Veterinary Record* **116**: 431-436.
- Morton, D.B. and Townsend, P (1992). Perioperative Care of the Pig. Invited paper published in the Proceedings of the *1st International Micropig Symposium* Volume 3.
- Morton, D.B. and Townsend, P (1995). Dealing with adverse effects and suffering during animal research In revised version of *Laboratory Animals - an introduction for experimenters*. (Ed. A.A. Tuffery). John Wiley and Sons Ltd, Chichester, UK. pp 215- 231.
- Morton, D.B. (1985). "Some Areas of Concern in the Care and Use of Experimental Animals" *First BVA Animal Welfare Foundation Symposium*. BVA, London.
- Morton, D.B. (1986). "The recognition of pain, distress and discomfort in small laboratory mammals and its assessment." *Second BVA Animal Welfare Foundation Symposium*. BVA, London.
- Morton, D.B. (1987). Epilogue : Summarization of Colloquium Highlights from an International Perspective. *J. Amer. Vet. Med. Ass.* **191**: 1292-1296.
- Morton, D.B. (1989). The Animals (Scientific Procedures) Act 1986. In: *Animal Welfare and the Law*. (Eds D.E. Blackman, P.N. Humphries and P. Todd) Cambridge University Press. Chapter 11, pp.195-219.

Morton, D.B. (1990). Adverse effects in animals and their relevance to refining scientific procedures. *ATLA* **18**: 29-39.

Morton, D.B. (1992). A fair press for animals. *New Scientist* 11th April 1992 No. **1816**. pp 28 - 30.

Morton, D.B. (1994). Enrichment techniques for rodents and rabbits. In: *Rodents and rabbits: current research issues*. (Eds. Niemi, S.M., Venable, J.S. & Guttman, H.N.). SCAW and WARDS, Washington.

Morton, D.B. (1994). The need to refine animal care and use. In: *Alternative Methods in Toxicology and the Life Sciences: Proceedings of the World Congress on Alternatives to Animal Use and the Life Sciences: Education, Research & Testing*. (Eds. Goldberg, Alan M. & LFM Zutphen). Mary Ann Liebert, New York. pp 35-41.

Morton, D.B. (1995). Advances in Refinement in Animal Experimentation Over the Past 25 Years. *ATLA* **23**: 812- 822.

Morton, D.B. (1995a). Practical ideas for refinement in animal experiments. *Proceedings of Animals in Science Conference Perspectives on their Use, Care and Welfare*. (Ed. N.E. Johnston). Monash University. pp 15 -167.

Morton, D.B. (1995b). Recognition and assessment of adverse effects in animals. *Proceedings of Animals in Science Conference Perspectives on their Use, Care and Welfare*. (Ed. N.E. Johnston). Monash University. pp 131-148.

Morton, D.B. (1995c). The post-operative care of small experimental animals and the assessment of pain by score sheets. *Proceedings of Animals in Science Conference Perspectives on their Use, Care and Welfare*. (Ed. N.E. Johnston). Monash University. pp 82-87.

Morton, D.B. (1997a). A scheme for the recognition and assessment of adverse effects. In: *Animal Alternatives, Welfare and Ethics*. (Eds. LFM van Zutphen and M. Balls). Elsevier Science, Amsterdam. pp. 235-241.

Morton, D.B. (1997b). Ethical and Refinement Aspects of Animal Experimentation. In: *Veterinary Vaccinology*. pp. 763 - 785. (Eds. Pastoret, P.-P., Blancou, J., Vannier, P. and Verscheuren, C.) Elsevier Science. pp. 853.

Morton, D.B. (1997c). Physiological and philosophical aspects of animal consciousness and self-awareness. In *Building healthy babies: the importance of the pre-natal period*. Supplementary papers on "Psychological and neurological development in the human and animal pre-born". Women and Children's Welfare Fund. Jedburgh, UK.

Morton, D.B. (1998a). The recognition of adverse effects on animals during experiments and its use in the implementation of refinement. *Proceedings of the Joint ANZCCART / NAEAC Conference on Ethical Approaches to Animal-based Science* (Eds. D. Mellor, M. Fisher and G. Sutherland), ANZCCART, Adelaide and Wellington. pp: 61-67.

Morton, D.B. (1998b). The use of score sheets in the implementation of humane end points. *Proceedings of the Joint ANZCCART / NAEAC Conference on Ethical Approaches to Animal-*

based Science (Eds. D. Mellor, M. Fisher and G.Sutherland), ANZCCART, Adelaide and Wellington. pp: 75-82.

Riley, V. (1981). Psychoneuroendocrine Influences on Immunocompetence and Neoplasia. *Science* **212**: 1100-1109.

Rose, M. (1987). *Biological definition of animals in experimental design. ACCART A role for the veterinary profession*. Available from ANZCCARTPO Box 19 Glen Osmond, South Australia 5064, Australia

Rose, M. (1994). Environmental factors likely to impact on an animal's well-being- an overview. In : *Improving the Well-being of Animals in the Research Environment*. (Eds. Baker, R.M., Jenkin, G. & Mellor, D.J.) ANZCCART, Adelaide. pp. 99-116.

Smith, J.A. and Boyd, K. (1991). (Eds.) *Lives in the balance. The Ethics of Using Animals in Biomedical Research*. The Report of the Working Party of the Institute of Medical Ethics. Oxford Univ. Press, UK. pp 352.

Soothill, J.S., Morton, D.B. and Ahmad, A. (1992). The HD50 (hypothermia-inducing dose 50): an alternative to the LD50 for measurement of bacterial virulence. *International Journal of Experimental Pathology* **73**: 95-98.

Stauffacher, M. (1992). Group housing and enrichment cages for breeding and fattening and laboratory rabbits. *Animal Welfare* **1**: 105-125.

Tuli, J. (1993). *Stress and parasitic infection in laboratory mice*. PhD Thesis, University of Birmingham, UK.

United Kingdom Co-ordinating Committee on Cancer Research (UKCCCR) Guidelines for the Welfare of Animals in Experimental Neoplasia (Second Edition). Workman, P., Twentyman, P., Balkwill, F., Balmain, A., Chaplin, D., Double, J., Embleton, J., Newell, D. Raymond, R., Stables, J., Stephens, T. and Wallace. J. *British Journal of Cancer* 1998, **77**: 1-10.

Universities Federation for Animal Welfare (UFAW) *Handbook on The Care & Management of Laboratory Animals* (1987). 6th Edition. Longman Scientific and Technical, London.

van den Heuvel, M.J. , Clark, D.G., Fielder, R.J., Koundakjian, P.P., Oliver, G.J.A., Pelling, D., Tomlinson, N.J. and Walker, A.P. (1990). The international validation of a fixed dose procedure as an alternative to the classical LD50 test. *Food Chemical Toxicology* **28**: 469-482.

van den Heuvel, M.J. , Dayan, A.D., & Shillaker. (1987). Evaluation of the BTS Approach to the Testing of Substances and Preparations for their Acute Toxicity. *Human Toxicology* **6**: 279-291.

van Zutphen, L.F.M., Baumans, V. and Beynen, A.C. (1993). *Principles of Laboratory Animal Science: A contribution to the Humane Use and Care of Animals and to the Quality of Experimental Results*. Elsevier, Amsterdam.

Wadham, J.J.B. (1996). *Recognition and reduction of adverse effects in research on rodents*. PhD Thesis, University of Birmingham, UK.

Wallace, J., Sanford, J., Smith, M.W. and Spencer, K.V. (1990). The assessment and control of

the severity of scientific procedures on laboratory animals. Report of the Laboratory Animal Science Association Working Party (assessment and control of severity). *Laboratory Animals* **24**: 97-130.

Wayneforth, H.B. & Flecknell, P.A. (1993). *Experimental and Surgical Technique in the rat*. 2nd Ed. Academic Press, London.

Wixson, Sally, K. (1994). Rabbits and Rodents" Anesthesia and Analgesics. In: *Research Animal Anesthesia, Analgesia and Surgery*. (Eds. Alison C. Smith and M. Michael Swindle). Scientists Center for Animal Welfare. SCAW, USA. pp. 59-92.

Wolfensohn, S. and Lloyd, M. (1994). *Handbook of laboratory animal management and welfare*. Oxford University Press.

