Revision of a Scale for Assessing the Severity of Live Animal Manipulations

Virginia M Williams[1], David J Mellor[2] and John Marbrook[3]

---

[1] School of Natural Sciences, Unitec, PB 12025, Pt Chevalier, Auckland, New Zealand.
vwilliams@xtra.co.nz
[2] Animal Welfare Science and Bioethics Centre, Massey University, Palmerston North, New Zealand
[3] 8 Brighton Rd, Parnell, Auckland, New Zealand

Abstract

In 1997, a severity scale to assess and record the level of welfare compromise to animals used in research, testing and teaching was introduced in New Zealand. Under this scale, the severity of procedures was expressed in terms of different categories of suffering based on numerous examples at the five levels outlined in a paper by Mellor and Reid (Mellor and Reid 1994). This paper reports on a review into the operation and effectiveness of that scale and the extent to which it fulfils the purposes for which it was devised. Key features of the scale are described, including its strengths and limitations, and comparisons with other scales operating internationally are made. Recommendations regarding modification of the scale based on this evaluation are outlined and key steps in its implementation are described.

Introduction

The New Zealand Animal Welfare Act (1999) requires that any person or institution wishing to manipulate live animals for the purposes of research, testing or teaching must operate under an approved Code of Ethical Conduct (CEC). The code requires that an Animal Ethics Committee (AEC) be set up so that all projects can be scrutinised and only approved projects may proceed. Although the Animal Welfare Act itself is relatively recent, this system has in fact been in operation since 1987 as a result of the process used to implement the Animals Protection (Codes of Ethical Conduct) Regulations 1987.

One of the functions of the AEC is to carry out a cost-benefit analysis of projects, with more 'invasive' or 'severe' manipulations requiring greater justification for the work in terms of the anticipated benefits that accrue from it. Applicants presenting protocols to AECs must therefore predict the level of animal welfare compromise expected as a result of the proposed manipulations, although that assessment may be corrected once the manipulation has taken place if the actual degree of compromise is found to have varied from that expected.

As detailed by Bayvel (2004), the data from such assessments are collected for each approved study, accumulated on an institutional basis and, after submission to the Ministry of Agriculture and Forestry (MAF), collated as annual national figures. These provide published details of the number of animals of different types which were exposed in New Zealand to manipulations of different severity levels on the

specified scale. Thus a more detailed picture of annual animal use in New Zealand is provided than would be from a grand total for animal use.

In assessing welfare compromise, New Zealand uses a five-point scale devised by Mellor and Reid (1994), which was introduced in 1997. Concerns have been raised, however, about a probable lack of consistency across institutions in the way the scale has been applied. This could have resulted in both under- and over-assessment of the impact of manipulations. A tendency for the collected statistics to be misrepresented has been dealt with in part by a new requirement in the collection of statistics by MAF which differentiates animal use according to whether it is for research, testing or teaching. Thus the relatively large numbers of animals that are used in testing required by legislation, some of which have a high degree of welfare compromise, are separated out within the total number of animals used. This confirms, for the concerned public, that experimental research or teaching rarely involves animals at high levels of "suffering", contrary to the charge that is often levelled at researchers by animal rights protestors.

It has also become apparent that the purposes of the scale are not well understood by some of those using it, with the result that it has, in a few cases, been applied by simply using the minimal definitions (no suffering, little suffering, moderate suffering, severe suffering and very severe suffering) found on the MAF statistics collection form. Others have merely used the examples provided in the full explanation provided by MAF (Anon [e] 2001) in a way that prevents the greater flexibility and judgement allowed under the five domains as intended in the original paper (Mellor and Reid 1994).

Concerns have also been raised about whether the use of animals in biotechnology is adequately covered within the current system, with the suggestion that separate categories might be more appropriate.

Accordingly, as part of a process of continuing improvement followed since the system was first introduced in 1987, MAF commissioned research to review the rationale, terminology and explanation of the scale and to compare it with others now used overseas, with the aim of ensuring clarity and minimising ambiguity in definitions and descriptions of purposes and differences between categories.

Purpose of grading the impact of manipulations

A severity scale for animal-based scientific manipulations is of significance to four interested parties: animal-based scientists, animal ethics committees, regulators and the public.

- **Animal-based scientists** need to assess the invasiveness of their manipulations as an integral part of seeking approval to undertake them. Evaluating invasiveness gives an indication of the harm that may be done to the animals. Conducting a harm-benefit analysis is a pivotal part of achieving approval for a proposed manipulation. This ensures that the potential adverse effects on the animals used are greatly outweighed by the benefits accrued from the work. This way of justifying the work is in compliance with the utilitarian ethical basis for using animal in research, testing and teaching.

- **Animal ethics committees** are required by the Animal Welfare Act and their institution's Code of Ethical Conduct to undertake a harm-benefit analysis for each application to manipulate animals in research, teaching and testing. This too relates to the Utilitarian ethical mode of assessing the justification for such animal use as proposed in applications to each committee. It also allows members of the AEC to help to ensure that the principle of refinement, or minimisation of harm, is fulfilled.

- **Regulators** require it because it confirms that such assessments are indeed being undertaken in compliance with the Animal Welfare Act, and this can be demonstrated in the public release of annual statistics showing the range of invasiveness of approved manipulations.

- **Interested members of the public** desire knowledge of this kind in order to be reassured that (1) excessively invasive manipulations are not being conducted in a high proportion of the animals, (2) very invasive manipulations are properly justified, and (3) within the full range of severity, the majority of manipulations have been at the bottom end of the range where the impacts are very low (benign) with few negative consequences.

Key features of the current scale

- **Five levels of severity are defined.** They are O, A, B, C and X, and range from no, or virtually no, impact (Grade O) to the most severe impact that one could imagine would ever be approved (Grade X). They are currently expressed in the Animal Welfare (Records and Statistics) Regulations 1999 in terms of severity of suffering, although this was not the mode of expression employed by Mellor and Reid (1994) in their original description of the scale. They referred to different levels of "ethical cost", not suffering.

- **The predicted severity of welfare impacts is assessed comprehensively.** This is achieved by reference to *five domains of potential welfare compromise* which relate to the

nutritional, environmental, health maintenance, behavioural and mental needs of animals. The final assessment of the severity of welfare compromise refers mainly to the *mental domain* because the overall welfare status of an animal is directly reflected in its affective or mental state. Nevertheless, consideration of the other four domains is imperative as it helps to ensure that the welfare status of the animals to be manipulated has been assessed thoroughly, in all of its dimensions.

- **The predicted severity of impact determines the required level of justification.** The different levels of justification were characterised by Mellor and Reid (1994). The greater the impact, i.e. the more severe the manipulation, the greater is the justification that is required before animal ethics committee approval can be given. Those justifications range from no requirement to demonstrate immediate or even long-term benefits for people or animals for Grade O manipulations, to the most exceptionally strong justifications in terms of benefits to people and/or animals for Grade X manipulations. Grade X should contribute to resolving a pressing need of great significance.

International comparisons

A paper that considered the possibility of international harmonisation of "pain and distress" classification systems (Purves 2000) highlighted the considerable international variation in the way these systems were used.

In the present critical review, the New Zealand system was compared with nine other current scales (Table 1), all of which were included in the Purves paper (2000) except one developed by New South Wales Agriculture in Australia (Anon [b] 2003). The Swedish system, which was legally abandoned in 1989 and only recently reinstated, was omitted.

Seven of the ten systems used classifications based only on the effect on the animal, with descriptors including discomfort, stress, pain, sorrow, harm, fear, suffering, severity degree and constraint. In addition, some of these included consideration of the duration of manipulation.

Of the other three, the Australian scale alone classifies types of procedures rather than effect on the animal, while the other two combine features of both systems: the New South Wales system outlines broad categories of intervention, differentiating between minor and major surgical and physiological challenges, while the US system, having separated out procedures causing little or no pain and distress, simply divides the remainder according to whether or not pain relief is given.

The Impact Scale

The demands for transparency by interested members of the public will not be met if, for whatever reason, the categorisation within the scale is not accurate or meaningful, the scale does not give a balanced view of overall animal use or the scale does not support the production of meaningful statistics. The present critical review confirmed the sound basis for categorisation set in the original paper by Mellor and Reid (1994), thus recognising that the total impact on an animal depends on more than the specific manipulation being carried out e.g. the competence of those carrying out the manipulation. However, the review also identified areas where greater clarity of purpose and more effective practical implementation could be achieved.

These included:
- Use of more appropriate and balanced descriptors of steps on the scale;
- Ensuring that the scale categories collectively cover the whole range of possible manipulations;
- Ensuring greater accuracy in categorisation through predicting impact on individual animals or groups within an experiment rather than giving one grade to a whole experiment;
- Ensuring greater accuracy of statistics by requiring that gradings reported to MAF reflect actual rather than predicted impact;
- Ensuring that category descriptors and examples are seen as guidelines rather than as prescriptive requirements, thus emphasising the importance of judgement when assessing individual cases;
- Determining whether the scale will adequately cover new developments such as those used in biotechnology;
- Ensuring that all factors with the potential to impact on animals are taken into consideration, for example the provision of enrichment, and special nursing techniques, as well as the use of analgesia;
-
-

Given the role of the classification system in following statistical trends in the distribution of manipulations from year to year, as well as the need to ease transition to a modified scale, there would be value in ensuring that any changes to the system are compatible with the previous scale.

The influence of terminology – "impact" vs "severity" and "suffering"

The Mellor and Reid (1994) paper discusses potential welfare compromise, ethical cost and the impact of procedures on animals. It outlines a rigorous method, based on functional criteria and informed judgement, to rank the severity of manipulations and assign them to specified categories. This has commonly become known as the "severity scale". In practical terms, on animal ethics application forms and on the statistics forms required to be sent back annually by institutions using animals to MAF, the five grades, each of which is explained in some complexity within the body of the Mellor and Reid (1994) paper, have been reduced to the following simple definitions – no suffering or virtually no suffering (0), little suffering (A), moderate suffering (B), severe suffering (C) and very severe suffering (X) (Anon 2004).

Given that one of the reasons for reviewing the scale was to assess whether it best meets the needs of all those who use it and the interested public, it was important to determine if misconceptions can arise simply from the naming of the categories. Both "severity" and "suffering" are words with negative connotations. While they may be appropriate when the degree of welfare compromise is high, it is perhaps misleading to label the whole grading system in this way. For instance, in the New Zealand context at least, animals which may be manipulated simply by a change in diet, thus falling into the minimal impact "0" category, are still categorised in terms of "suffering", albeit "little or no suffering". The concept of suffering has a particular meaning when used in the assessment of welfare, including as it does varying degrees and combinations of anxiety, fear, pain and distress. In a wider "lay" context, however, it is an emotive word and one that can have very different meanings to different people. The word "severity" in relation to the grading scale likewise sets a negative context.

The recommendation to replace both "severity" and "suffering" with the more neutral word "impact" is not in any way meant to imply that some animals do not suffer when used for research, testing and teaching. One of the impacts on animals may indeed be suffering, but having the words "severity" and "suffering" to describe the scale is clearly not precise when they are already used as descriptors for categories **within** the scale. Rather, the change is an acknowledgement that while there is always an impact, suffering does not always occur. "Impact" covers all effects that any procedure may have on an animal, while "suffering" does not. Thus the words "suffering" and "severity" may be used within categories of the "impact" scale, but do not define them.

Number of categories

The need to classify the wide range and diverse characteristics of manipulations in animal-based research, teaching and testing into categories has been well established. The present system aims to

cover all possible procedures that might be carried out on animals used in research, testing and teaching, but it is suggested here that classification systems should also include manipulations, referred to by Banner et al (1995), which would not be allowed under any circumstances.

The comparison of scales used internationally showed that most used between three and five categories, and reviews have raised suggestions that overcomplication may result from too many categories (Smith and Jennings 2004). While New Zealand's current 5-category scale starts at the lowest level where the manipulations involve no, or very little, invasiveness or severity and progresses to the highest such severity rating that would only very rarely, and for the most compelling of reasons, be approved, it is recommended that a sixth category of manipulations be added to include manipulations, which, while quite possible, are of a severity that is unacceptable. It is proposed that this grading be designated Z. Modified thus, the impact scale would then encompass the full range of possible manipulations. This puts the range in true perspective as it indicates that the small number of manipulations that are graded X, while of high impact, are not as high as is possible. Without this extra grading, the impression is conveyed that X manipulations are "as bad as it gets". If this were to be adopted, the annual national score for this sixth category (Z) would therefore always be zero.

Segregation of manipulation categories within an experiment

The Animal Welfare Act 1999 lacks some clarity in how animals used in a single experiment are categorised. Quite frequently there is a marked difference between groups within an experiment. For example, severe impact groups may be compared with negative controls. This can, and has been, interpreted in two ways. Either the whole experiment is categorised according to the most severe category or groups are segregated and individual groups categorised separately. In collecting statistics, it is essential to be as accurate as possible but as importantly, to be consistent.

Greater accuracy is obtained if scoring is carried out separately on segregated groups or even individuals. Untreated controls, for example, would not be counted with other groups of treated animals thus inflating the number of animals receiving high impact treatments. This is particularly important in experiments in which animals are compared in groups with large numbers, but also, for example, where a series of groups of animals may be set up in which the effect of a substance is measured as a dose response. The groups may vary as a graded series from nil to high impact, requiring a measure of judgement on behalf of both researchers and animal ethics committees.

Reporting of actual, rather than predicted, impacts

The prediction of impact is necessarily part of an application for ethics approval prior to animal manipulations taking place. In many cases, particularly those where the researcher has extensive experience with a particular manipulation, those predictions will be accurate. However, this is not always the case. There may be instances where a manipulation results in a lesser, or indeed a greater, than predicted impact.

It is important in terms of accuracy of the statistics that the gradings listed in the annual returns to MAF reflect not the predicted impact but the actual impact. This requires that those manipulating animals keep accurate records of the actual impact during the experimental or testing period for subsequent reporting, a practice that is not universally carried out at present. This practice not only enables greater accuracy of reported statistics, but should also allow more accurate predictions of similar manipulations in subsequent experiments. This is important for both the applicants and the AEC members, for whom such information informs the monitoring process which is part of their remit.

Category descriptors as guidelines

It is important that the grading system not be seen as inflexible. It is not possible to produce a workable scale that specifically takes account of all possible manipulations in all possible species. This necessitates the use of various qualitative magnitude terms such as "short term", "rapid", "protracted " etc where what is meant very much depends on a number of factors, such as species and type of manipulation. This means that any assessment or prediction of impact will require a measure of judgement on the part of both applicant and AEC. It also underlines the importance of regrading manipulations upon completion of the study (as noted above), not only to provide accurate statistics, but also to widen the knowledge base of the effect of such interventions.

Animal manipulations in biotechnology

There is a public perception that the accelerating development of biotechnology has provided completely new challenges with respect to the ethics of using animals in research. As a result there has been an active societal debate on the acceptability of biotechnology, particularly in relation to specific processes such as genetic manipulation.

In the context of this discussion, biotechnology can be defined as *the use of techniques such as those of 'molecular biology' to effect functional changes at the tissue and/ or whole-animal level.* Changes may be temporary or may become a heritable trait.

Leaving aside the wider ethical question of the acceptability of biotechnological applications in general, the concern here is with the more specific question of whether such manipulations are acceptable in

animal welfare terms. That is, whether they meet currently appropriate and generally accepted ethical criteria for animal use in research and can be assessed with the expertise currently available on ethics committees as set up in New Zealand.

Examples of such manipulations include:

(i) Manipulating the gamete (ovum, spermatozoon) or embryo;
(ii) Addition to or deletion from the genome;
(iii) The cloning of animals;
(iv) The repopulation of tissues or creation of whole animals with stem cells;
(v) The secretion of proteins originating from other organisms;
(vi) Xenotransplantation.

The following observations on some of these examples can be made:

- At the present time, manipulating the ovum, during cloning for example, does not require ethical approval in New Zealand because an ovum is not considered to be an "animal" under the Animal Welfare Act 1999. Moreover, the associated whole-animal manipulations are identical to, and standard for, *in vitro* reproductive manipulations which are now routine and have received AEC approval for some time. However, the potential for such manipulations to affect the phenotype of resultant offspring has led to a review currently being conducted by the National Animal Ethics Advisory Committee of the exclusion of ova manipulation from the need for ethical approval. (See final bullet point below for monitoring of animals with new phenotypes).
- Changing the genome of animals with the use of mutagens has been a process carried out for many years, although now largely replaced by biotechnology. Screening for mutations is already carried out in investigations of unwanted side effects with candidate anti-cancer drugs under AEC approval.
- Stem cell biology has been studied for many years. Altering stem cells prior to transfer is carried out *in vitro* and therefore presents no novel ethical considerations for AECs, which focus on the welfare implications of the introduction of such cells into whole animals.
- The xenotransplantation debate mainly involves animal-to-human transfers. Notwithstanding this, and acknowledging that obtaining pathogen-free supply animals does have special welfare considerations, any associated consequences can be well managed by current AEC procedures which have applied to animal-to-animal transfers conducted for many years.
- Breeding offspring from founder animals may appear to be novel, but the processes have been long carried out in mice in producing inbred lines and recombinant inbred lines.

Clearly, the presence of new genes, for example, may produce different phenotypes depending on the gene background and whether the animal is heterozygous or homozygous with respect to the new gene, which may have negative animal welfare impacts. Again, these can be well managed by current AEC procedures.

In summary, whether animals used in biotechnology are the providers or carriers of manipulated genetic material, the resulting offspring and even the subsequent generations, related animal welfare considerations can provide some new challenges in terms of rate and scale of change. The vast majority of the manipulations involved are variations on current technologies which have provided few problems during AEC assessment. Although extra care will be needed to monitor the welfare of animals having phenotypes with unknown welfare implications, this requirement falls within the current remit of animal ethics committees. This is supported by a paper (de Cock Buning 2004) which concluded that as long as the welfare of animals at all stages of a biotechnological procedure is assessed by an animal ethics committee, the wider discussion surrounding these processes should be addressed in a different forum.

Factors impacting on animals in research, teaching and testing

The five domains of compromise which relate to the nutritional, environmental, health (injury, disease), behavioural and mental needs of animals as suggested by Mellor and Reid (1994) provide an appropriate basis for classification as they allow an holistic approach to assessing impact, an approach that has been endorsed by other reviews of such categorisations (Orlans 2000; Smith and Jennings 2004). The current scale goes through physical impacts first and then addresses the mental state of the animal. However, it is now suggested that a provisional score with respect to mental state be established first as the ultimate measure of impact, with the other four as contributors to that ultimate measure being double-checked to ensure no factor has been missed, nor the impact with regard to mental state over or underestimated.

Alongside and contributing to the "domain" assessment, both intrinsic (animal) and extrinsic (environmental, procedural) factors which will influence the impact on the animal need to be taken into account. These include:

Animal factors
- Species/type/breed differences - marked variation between and within species can occur in terms of, for example, responses to human presence, fear responses and temperament. A detailed understanding of unique and common features of the biology and behaviour of each species studied, including life stage differences, is essential in assessing welfare

impact. Moreover, an adequate application of the scale may necessitate consultation with others with greater expertise in the species of interest.

- Size – the impact of factors such as needle gauge, injection volumes and extent and complexity of surgical manipulations relative to the size of the manipulated animal needs to be accommodated.

Environment

- Provision for social and behavioural needs, including the enrichment of environments, should be taken into account in the categorisation of procedures. Likewise the influence of physical environmental features, including temperature, light, sound, air quality, vibration, space, comfort of ground/floor surfaces, safety and other aspects need to be assessed. Also, the impact of animal attendants and investigators who are part of the animals' environment should be considered.

Procedures

- Types of manipulation – the welfare impact of each manipulation relates to both the magnitude of negative effects and their duration, as well as to the interval between successive manipulation e.g. for recovery. The five domains of potential welfare compromise (Mellor and Reid, 1994) provide a comprehensive way of assessing impact. Any guidelines relating to specific procedures should be indicative, not prescriptive, requiring the exercise of good judgement when being applied.

The final grading will depend on other factors as well, including:

- Operator skill – there may be need to grade skilled operators and trainees differently, for example. This emphasises the need for some flexibility in grading procedures.
- Harm minimisation – the extent of application of the Three Rs to proposed manipulations will affect the final grading.

Summary

1. The basis of the system in the Mellor and Reid (1994) paper remains appropriate.
2. The name of the categorisation system should be the "impact scale".
3. The current 5-point system should be enlarged by the addition of a sixth category which includes procedures that must not be carried out under any circumstances. To minimise confusion in the transition from one scale to another, it is suggested that the original categories retain their symbols of 0, A, B, C and X, while the sixth should be labelled Z.

4. To ensure greater accuracy, where individuals or groups within an experiment are likely to experience significantly different impacts, they should be graded on an individual or group basis, rather than a whole experiment basis as was originally suggested by Mellor and Reid (1994).

5. Animal manipulations in biotechnology are adequately covered by the present system and the addition of special categories is not necessary.

6. An exhaustive list of manipulations with recommended gradings is not advisable, both because it will inevitably be incomplete and because it tends to be viewed in a rigid manner. Implementation of all the recommended features will require judgement to be exercised by scientists and AEC members alike.

7. A requirement should be made for the predicted impacts of manipulations to be reconsidered at the end of each study, to ensure that the actual impacts, as assessed by researchers, animal-care staff, animal welfare officers and AECs, are supplied in the annual statistical returns to MAF.

Acknowledgements

References

Anon (a) (1991). Categories of invasiveness in animal experiments. Canadian Council on Animal Care. http://www.ccac.ca/english/gui_pol/policies/CATEG.HTM Accessed 14/11/04.

Anon (b) (2003). Form L Animal Use Statistics, New South Wales Agriculture. http://www.animalethics.org.au/reader/arrp-forms/form-l2003.doc?MIvalObj=19990&doctype=document&MItypeObj=application/msword&ext=.doc Accessed 11/01/05.

Anon (c) (2004). Classification prospective des experiences sur animaux selon leur degre de gravite (categories de contrainte), Swiss Federal Veterinary Office. http://www.bvet.admin.ch/tierschutz/00237/00833/index.html?lang=fr&download=00423_fr.pdf Accessed 11/01/05.

Anon (d) (2004). Taking animal welfare seriously: minimizing pain and distress in research animals. Humane Society of the United States. http://www.hsus.org.ace/Article_Printer_Friendly?Content_ID=11425 Accessed 12/07/04.

Anon (e) (2004). Animal use statistics. Wellington, Ministry of Agriculture and Forestry.

Banner M, Bulfield G, Clark S, Gormally L, Hignett P, Kimball H, Milburn C, Moffitt J. Report of the Committee to Consider the Ethical Implications of Emerging Technologies in the Breeding of Farm Animals. Ministry of Agriculture, Fisheries and Food (MAFF), London, 1995

Bayvel A C D (2004) Animal use Statistics: The New Zealand and Australian Experience and Perspective, ATLA 32 (Supplement) 287 - 292

de Cock Buning, T. (2004). "Practical difficulties in balancing harms and benefots in the modern use of laboratory animals: biotechnology." ATLA 32 (Supplement 1): 459-463.

Mellor, D. and C. S. W. Reid (1994). Concepts of animal well-being and predicting the impact of procedures on experimental animals. In *Improving the Well-being of Animals in the Research Environment (pp 3-18).* Sydney, ANZCCART.

Orlans, F. B. (2000). Practical applications of animal harm scales: international perspectives. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (pp 1049-1056). New Orleans, Elsevier Science.

Orlans, F. B. (2000). Public policies on assessing and reporting degrees of animal harm: international perspectives. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (pp 1075-1082). New Orleans, Elsevier Science.

Purves, K. E. (2000). Global harmonisation of pain and distress classification systems: current analysis and opportunity for change. Orlans, F. B. (2000). In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (pp 1057-1067). New Orleans, Elsevier Science.

Smith, J. and M. Jennings (2004). Categorising the severity of scientific procedures on animals, The Boyd Group and the RSPCA.