

REPORTING THE SEVERITY OF ANIMAL PROCEDURES RETROSPECTIVELY

**Report of a LASA/APC Pilot Study to assess the feasibility of
collecting and reporting data on the severity of adverse effects
caused to animals used in procedures regulated under the
Animals (Scientific Procedures) Act 1986**

December 2005

Chairman of the Working Group:

David Smith (President, LASA)

Members of the Working Group:

David Everett (Covance)
Paul Finnemore (Oxford University)
Gill Fleetwood (GlaxoSmithKline)
Kathleen Mathers (MRC NIMR)
Peter Pearce (DSTL)
Ghislaine Poirier (GlaxoSmithKline)
Sally Robinson (AstraZeneca)
Clare Stanford (University College London)
Dominic Wells (Imperial College)
Sara Wells (MRC Harwell)

Observers:

Ken Applebee (King's College London)
Bob Kemp (APC Working Group on Suffering and Severity)
Peter Thornton (Animals Scientific Procedures Inspectorate)

Rapporteur:

Jane A. Smith

Executive Summary

- 1** In response to recommendations from a number of expert bodies, a working group was convened under the auspices of Laboratory Animal Science Association (LASA) and the Suffering and Severity Working Group of the Animal Procedures Committee (APC), to assess the feasibility of collecting and reporting data on the severity of adverse effects experienced by animals used in regulated scientific procedures.
- 2** The working group represented nine establishments drawn from industry/ pharmaceutical organisations (3), large universities (3) and major government research institutes (3). The roles of project licence holder, personal licensee, Named Veterinary Surgeon and Home Office liaison officer were represented in the group.
- 3** A member of the APC Working Group on Suffering and Severity and a Named Animal Care and Welfare Officer (NACWO) also participated. A Home Office inspector attended all meetings as an observer.
- 4** Working within the terms of reference provided by the APC, the group sought to devise a method of providing information about suffering and severity actually experienced by individual animals that could be published in an annual report and be used to refine future experiments. The process should have neutral or the least additional regulatory impact in terms of resource.
- 5** In order to engage more widely with project licence holders, a questionnaire was used to establish current practice in the nine establishments. Analysis of 168 responses showed that 9 out of 10 licence holders whose work involves moderate and/or substantial protocols always or sometimes make records of adverse effects actually experienced by the animals. The biggest difficulty they foresee is the ability to record this information in a way that will facilitate annual Returns. Their major concern is the resource burden and its possible impact on time for welfare and science.
- 6** Drawing on the questionnaire findings, experience of working group members and, in light of reporting models used in other countries (e.g. Switzerland), a range of initial options for reporting severity were identified and discussed.
- 7** Alongside the identification of options, working group members brought to the discussions examples of protocols from their own establishments, demonstrating adverse effects from the procedure and/or its outcome. The protocols involved a variety of laboratory mammal species, including non-rodents; and had the potential to cause effects spanning severity categories mild, moderate and substantial.
- 8** The first option considered was reporting number of animals used by protocol severity limit, which is an improvement on the current publication of severity bands, but is not reflective of actual adverse effects. It would greatly over-estimate actual severity and provide seriously misleading information.
- 9** Reporting using a single code (mild, moderate, substantial or unclassified) to indicate maximum actual severity was considered and found to be meaningful for simple protocols, but was not capable of adequately capturing the effects of more complex, longer-term procedures that generate more variable severity profiles over time.
- 10** Combining intensity and duration into single severity codes, as in the Swiss system, would likely cause significant difficulties in interpreting the data reported. Increasing the number of categories, so as to avoid having to combine into a single code more

substantial (but relatively short-term) effects and milder (but longer-term) effects, may be a possible solution.

- 11 In order to capture the overall experience of the animal for the duration of a procedure, an intensity-duration severity grid, in which the two parameters are considered independently, was developed. This grid works well with procedures that have relatively simple severity profiles but again is less successful in providing a meaningful reflection of severity in more complex, longer-term procedures.
- 12 Reporting using two intensity-duration grids, to indicate (i) maximum severity and (ii) severity over the remainder of the procedure, has the potential to provide a representative picture of intensity and duration of severity over a wide range of different procedures. Early feedback from licence holders suggests that the system is understandable, intuitive to apply and acceptable in terms of its capacity to portray the severity of adverse effects in more complex procedures. The number of potential codes is significantly fewer than first envisaged and appropriate coding will frequently be predictable in advance, allowing exception reporting to be employed.
- 13 Options for Returning the retrospective data (and resulting statistics) were identified, but no simple solution was identified. Providing these data as an additional report to the 'prospective' Returns does not allow cross-reference to other parts of the published *Statistics* and would constitute an additional burden on licence holders and establishments. An integrated process would be ideal, but would require substantial modification of the current Returns system so that all the data are compatible, and also simplification to make the process manageable and acceptable.
- 14 Any scheme for retrospective reporting needs to be supported by detailed guidance, including a catalogue of worked examples covering the full range of species and a wide variety of regulated procedures. A glossary of severity codes for a range of protocols and outcomes, agreed by project licence holders, is essential. Such guidance would also be valuable in explaining the data when they are presented in the public domain. Worked examples in this report provide a useful start.
- 15 All working group members consider the introduction of a retrospective severity assessment process to be beneficial, but are mindful of the additional bureaucratic burden that this would bring. The working group is collecting estimates of resources required for introduction of retrospective reporting of severity, and early indications are that for some establishments the burdens will be considerable. This would only be acceptable if matched by a similar reduction in other activities associated with the operation of A(SP)A. A number of such activities have been identified during the course of this project, and could be taken forward through the Government's Better Regulated Task Force (adopting the Netherlands' "one-in, one-out" principle).
- 16 An extension of this work is proposed to:
 - consider simplification of the proposed reporting scheme;
 - assess its appropriateness across a wider range of establishments;
 - evaluate the consistency of response to the scoring system;
 - identify a process for developing a glossary of severity codes for a range of protocols and outcomes;
 - clarify a number of unresolved issues in the application of the coding scheme;
 - conduct a detailed impact assessment; and
 - assess the value of the scheme in generating data that can assist public understanding of the severity of effects caused to animals used in regulated scientific procedures.

Contents

	Page
1 Introduction	1
1.1 General background to the study	1
1.2 Events leading to LASA's involvement	1
1.3 Establishing the LASA/APC Working Group	3
2 Aims of the study	3
2.1 General terms of reference from the APC	3
2.2 Specific objectives and scope of the LASA/APC pilot study	4
3 Method of working	4
3.1 Questionnaire to project licence holders	4
3.2 Exploring possible methods for retrospective reporting of severity	5
3.3 Outputs	5
4 Questionnaire findings	5
4.1 Current practice in recording adverse effects	5
4.2 Ease of Returning information currently collected by severity category	6
4.3 Ease of recording adverse effects by severity category	6
4.4 General concern	7
4.5 Capacity to predict severity in advance of a study, then to report 'exceptions'	7
4.6 Views on subdivision of the 'moderate' severity category	7
4.7 Ways of improving the annual Returns procedure	7
5 Designing a system for retrospective reporting of severity	8
5.1 Assessing and recording <i>cf.</i> reporting adverse effects and their severity	8
5.2 Characteristics of a system for retrospective reporting of severity	8
6 Evaluating possible models for retrospective reporting	9
6.1 Reporting number of animals used by protocol severity limit	9
6.2 Reporting using a single code for each animal to indicate maximum actual severity	10
6.3 Reporting using an intensity-duration severity grid, to indicate either maximum or overall actual severity	13
6.4 Reporting using two intensity-duration severity grids, to indicate maximum severity and severity over the remainder of the procedure	14
6.5 Further development of reporting system	16
7 Issues for further consideration/resolution	16
7.1 Should there be a 'no effect' category?	17
7.2 Need for further guidance and examples on how to assign codes	17
7.3 Further developing consistency of approach in retrospective reporting	20
8 Options for Returning data collected	20
8.1 Prospective Returns with additional severity data, or retrospective Returns incorporating severity data?	21
8.2 Routine annual Return of severity information for all project licences, or selective/exemplar reporting?	21
9 Resource implications of a retrospective severity reporting process	22
10 Potential benefits of retrospective reporting of severity	23
11 Suggestions for further work	24

Appendices

	Page
1 Summary of responses to the questionnaire to project licence holders	<i>Separate document</i>
2 Examples considered by the Working Group	26
3 Degrees of severity categories, used for retrospective reporting of actual adverse effects in Switzerland	32
4 Suggested opportunities to reduce and/or simplify processes associated with A(SP)A if retrospective severity reporting is introduced	33

1 Introduction

1.1 General background to the study

1.1.1 At present, very little information pertaining to the severity of adverse effects¹ of scientific procedures on animals is made publicly available:

- Until recently, the only source of such information has been an unnumbered Table appearing in the annual *Home Office Statistics of Scientific Procedures on Animals*,² listing the number of project licences in force, granted and revoked during the relevant year by their overall 'severity band' (a prospective assessment of the likely experience of the 'average' animal within the project concerned), categorised as mild, moderate, substantial and unclassified³.
- Now, additional information is provided by the publication of abstracts of new project licences on the Home Office web-site. Some, but not all, of the abstracts published since January 2005 contain narrative descriptions of the severity of the anticipated effects of the licensed work on the animals involved⁴.

1.1.2 Both of these sources report *predictions* of *potential* adverse effects, not *retrospective* assessments of the *actual* adverse effects caused in practice; and so neither indicates the degree of pain, suffering, distress or lasting harm⁵ actually experienced by the animals involved in the projects. Moreover, the former source (information on the severity bands of project licences) only provides information on the potential adverse effects caused to the 'average' animal involved in a given project, and therefore gives no indication of the severity of effects likely to be caused to individual animals, nor of the anticipated effects of the different kinds of protocol used in the project.

1.1.3 Over the past few years, a number of expert committees and working groups have lamented the lack of publicly available information on the scale and degree of pain, suffering, distress and lasting harm caused to laboratory animals in Britain, and have recommended that retrospective information about actual adverse effects of regulated procedures on animals be collected and published. Box 1, overleaf, lists relevant recent reports and their recommendations. The issue of retrospective reporting of severity is also under consideration across Europe, within a more general review of EU statistical reporting requirements, which is being conducted as part of the current process of revision of EU Directive 86/609.

1.2 Events leading to LASA's involvement

1.2.1 In its response to the House of Lords' recommendation that "serious effort should be made to provide better statistics on animal suffering" (see Box 1), the Government agreed that "provision of further and better information on the life experience of each animal used in scientific procedures would be desirable" and that this "could serve to reassure the general public"; but that, "the difficulty of devising a method of capturing this information should not be underestimated"⁶.

¹ The phrase 'adverse effects' refers to all the adverse effects caused to animals by a protocol, encompassing effects that are anticipated (and therefore listed in the relevant 19b) and those which are unanticipated; also effects that are 'required' for the purpose of the protocol (e.g. induction of disease) as well as unwanted side effects.

² See page 86 in the *Statistics for 2003: Home Office (2004). Statistics of Scientific Procedures on Living Animals, Great Britain 2003*. Command 6291. TSO: London. <http://www.homeoffice.gov.uk/docs/animalstats.html>

³ For further information, see p. 32 in Home Office (2000) *Guidance on the Operation of the Animals (Scientific Procedures) Act HC 321*. TSO: London. www.official-documents.co.uk/document/hoc/321/321.htm

⁴ See <http://www.homeoffice.gov.uk/comrace/animals/index.html>

⁵ The terminology used in Section 2(1) of the Animals (Scientific Procedures) Act 1986, in defining a "regulated procedure".

⁶ Home Office (2003). *The Government reply to the report of the House of Lords Select Committee on Animals in Scientific Procedures. Session 2001-2002 HL 150-1*. TSO: London. <http://www.officialdocuments.co.uk/document/cm57/5729/5729.pdf>

Box 1: Recent recommendations on retrospective reporting of severity

2002. House of Lords Select Committee on Animals in Scientific Proceduresⁱ

Recommendation 31 of the Select Committee's report asserts that, "Serious efforts should be made to provide better statistics on animal suffering. The Home Office Inspectorate should develop or approve a 'scoring system' for animal suffering which could be operated by Named Animal Care and Welfare Officers and Named Veterinary Surgeons, and used to provide data for the *Statistics*."

2003. APC review of cost-benefit assessment in the use of animals in researchⁱⁱ

The APC concluded that, "Overall severity bands are inadequate both for purposes of cost-benefit assessment and providing public information about severity. On these grounds, we doubt the value of assigning overall severity bands for projects, and invite the Home Office to consider abandoning their use and putting in place a new system for public information purposes".

2004. Boyd Group/RSPCA report of discussions on categorising severityⁱⁱⁱ

Three focus groups were convened to consider the whole area of severity assessment and categorisation. These involved: (i) NVSs and NACWOs; (ii) animal protection and welfare group representatives; and (iii) project and personal licence holders. The consensus of all three groups was that, of several options, *retrospective reporting of the severity of adverse effects actually experienced by the animals* would provide the most pertinent statistical information about severity; but both the NVS/NACWO and licensees' groups were divided on whether the benefits of providing this information would justify the effort required to record it (see paragraph 1.2.3 for below further detail).

2005. Nuffield Council on Bioethics working party report on the ethics of research involving animals^{iv}

This report, which became available only towards the end of the pilot study, concluded that, "Information about the suffering that animals involved in procedures experience in practice is unsatisfactory. We recommend that the Home Office should make retrospective information about the level of suffering involved during procedures publicly available.... We recognise that the system of collecting data about numbers of animals used in research is very complex and care needs to be taken to avoid making existing administrative processes more onerous. Nevertheless, we think it highly desirable to present clearer information about how many animals of a particular species experience pain, suffering and distress, to what degree, and for how long. We therefore recommend that the *Statistics* be revised to provide this information..."

-
- i House of Lords (2002). *Report of the Select Committee on Animals in Scientific Procedures*. HL Paper 150-I. TSO: London. Available at <http://www.publications.parliament.uk/pa/ld/ldanimal.htm>
- ii Animal Procedures Committee (2003). *Review of cost-benefit assessment in the use of animals in research*. Home Office, Communications Directorate: London. <http://www.apc.gov.uk>
- iii Boyd Group/RSPCA (2004). *Categorising the severity of scientific procedures on animals: summary and reports from three round-table discussions*. Published by the RSPCA Research Animals Department. <http://www.boyd-group.demon.co.uk>
- iv Nuffield Council on Bioethics (2005). *The ethics of research involving animals*. Nuffield Council on Bioethics: London <http://www.nuffieldbioethics.org>

1.2.2 Further to this written response, the Home Office Minister then asked the Animal Procedures Committee (APC), to conduct "a small pilot study on the lines of that recommended by the House of Lords Select Committee, to establish what data might feasibly be captured on animal suffering", as part of a wider review of the content and format of the published Home Office *Statistics*. In view of the complexity of this task, and because, later, the Minister also asked the APC to take on "the development of practical recommendations about a new system of severity limits and bands", a new APC Working Group was convened towards the end of 2004 especially for this purpose (the 'Suffering and Severity Working Group', chaired by Professor Dawn Oliver).

1.2.3 In the meantime, LASA, having participated in the Boyd Group/RSPCA project, recognised a likely impasse in debates on the future of retrospective severity assessment, owing to concerns about the potential bureaucratic burden on both individuals and establishments. For example:

- The Boyd Group/RSPCA study found that "in both the NVS/NACWO and licensees' [focus] groups, opinion was divided on whether the benefits of providing [retrospective severity] information would justify the effort required to record it. For example, licensees whose projects involve using large

numbers of animals, and/or whose animal facilities hold large numbers of animals, were concerned that collecting these data would be overly burdensome. In the NVS/NACWO group there was concern that animal care staff need to preserve as much time as possible to check and tend to the animals in their care, rather than fill in forms for statistical purposes."

- In his letter to the APC, the Home Office Minister (at the time Bob Ainsworth MP) made clear that, in relation to retrospective reporting of severity, he was "particularly concerned about the value of what could be obtained, and the cost in terms of the additional work that might be involved for the scientific community".

1.2.5 In light of such concerns, and in order to help inform further discussions, LASA therefore proposed a feasibility study, which would:

- (i) appraise method(s) for recording and reporting actual severity; and
- (ii) provide more concrete evaluation of the potential benefits and difficulties involved, including likely resource impacts.

1.2.6 When it later became apparent that the APC had already been asked by the Minister to carry out a similar piece of work, LASA contacted the APC and, following discussions, it was agreed that LASA and the APC would work together on the pilot study and that the APC Working Group on Suffering and Severity would use the resulting report to inform its subsequent work.

1.3 Establishing the LASA/APC Working Group

The Working Group was established by its Chair, Dr David Smith (Safety Assessment, AstraZeneca and President of LASA), in consultation with the APC Working Group on Suffering and Severity. Working Group members were chosen to represent a range of different types of designated establishment, roles under the Animals (Scientific Procedures) Act 1986 [ASPA], regulated scientific procedures and species of animals:

- Three large industry/pharmaceutical establishments, three large universities, and three major government research institutes were represented.
- Representatives included project licence holders and personal licensees, Named Veterinary Surgeons [NVS] and Home Office liaison officers.
- A member of the APC Working Group on Suffering and Severity and a Named Animal Care and Welfare Officer [NACWO] also participated.
- A Home Office inspector attended all the meetings as an Observer.

2 Aims of the study

2.1 General terms of reference from the APC

In light of the Minister's letter requesting that the APC carry out a pilot study (see 1.2.2) above, the APC listed the following terms of reference for the project:

"To devise the most effective systems of retrospective assessment of suffering and severity that will achieve the following goals:

- Provide information about suffering and severity actually experienced by the animals used in a particular project that can be published in an annual publication of information and statistics;

- Provide information that will enable individual establishments and others to refine future prospective assessments; and.
- Have “neutral, or the least additional regulatory impact”.

2.2 Specific objectives and scope of the LASA/APC pilot study

2.2.1 In light of the above terms of reference, it was agreed that the overall aim would be:

"To assess the feasibility of conducting retrospective severity assessments using the framework of the 1986 Act"

2.2.2 Within this general aim, more specific objectives were:

1. To identify a representative sample of establishments conducting different types of scientific work on laboratory animals.
2. To review their existing processes for recording the actual severity of adverse effects caused to animals in scientific procedures.
3. To propose a simple process that would enable recording and reporting of actual severity.
4. To evaluate the potential impacts of conducting such retrospective assessments in practice:
 - to identify benefits and difficulties in practice and/or issues for further consideration;
 - to assess resource (including time) impacts.
5. To consider and propose options for Returning (to the Home Office) the data recorded and thereby to contribute to the annual publication of statistics and other information.
6. As an add-on, in light of the APC's recent recommendation⁷, to consider possible sub-division of the moderate severity category.

2.2.3 The study was carried out within the framework of regulation of severity currently used by the Home Office in administering the 1986 Act. Therefore, options for reporting actual severity were considered in the context of the present system of categorisation of severity of adverse effect⁸ (mild, moderate, and substantial⁹).

3 Method of working

3.1 Questionnaire to project licence holders

3.1.1 Members of the Working Group agreed to carry out the pilot study with open minds, setting aside any pre-conceived ideas about whether, and if so how, retrospective reporting of severity might be achieved in practice. Further to this position, and bearing in mind the likely resource implications of any such retrospective reporting system, it was agreed that it was vitally important that, from the outset, the study be

⁷ Animal Procedures Committee (2003). *Review of cost-benefit assessment in the use of animals in research*. Home Office, Communications Directorate: London. <http://www.apc.gov.uk>

⁸ The description of these categories in Home Office *Guidance on the Act* (Home Office 2000, p.32, *op cit.*) is brief. The present study is concerned with the development of a *method for reporting* judgements about severity, as currently made by licence holders. The Working Group recognises that there is a need for further guidance on the application of the categories in practice (see Section 7.2 below), and notes that the whole system of severity limits and bands, and their classification, is currently under review by an APC sub-group on (see Section 1.2.2 above).

⁹ The system also includes an 'unclassified' category, for protocols performed entirely under general anaesthesia, from which the animal does not recover consciousness; including preparation and use of decerebrate animals. Home Office (2000) *op cit.*

informed by the experiences and ideas of the people who would bear responsibility for collecting and Returning the data – that is, project licence holders.

- 3.1.2 To assist in this, the Working Group, at its first meeting, began to draft a questionnaire for project licence holders. The draft was piloted with small groups of licensees, refined on the basis of their comments, and then circulated to all project licence holders in nine designated establishments represented in the Working Group. The questionnaire is attached at Appendix 1, together with a summary of responses.

3.2 Exploring possible methods for retrospective reporting of severity

3.2.1 *Initial identification of models for reporting*

Drawing on the questionnaire findings, experiences of Working Group members, and in light of reporting models used in other countries, a range of initial options for reporting severity were identified and discussed.

3.2.2 *Examination of actual data*

Alongside the identification of options, Working Group members brought to the discussions examples of protocols from their own establishments. These protocols involved a variety of laboratory mammal species, including non-rodents; and, between them, had the potential to cause effects spanning all three severity categories. In each case, participants explored (i) how these effects might best be reported and (ii) where any difficulties lay.

3.2.3 *Refinement of initial ideas*

By looking at examples in light of findings from the questionnaire, the Working Group appraised the pros and cons of various options for reporting. This prompted early rejection of some possible models and, later, refinement of a preferred model, so as to accommodate as wide a range of experimental scenarios and adverse effect profiles as possible. This iterative process also enabled the Group to draw up a list of unresolved issues on which specific guidance would be needed.

3.3 Outputs

It follows that there are two main outputs from this pilot study:

- (i) project licence holders' responses to the questionnaire; and
- (ii) the Working Group's consideration of sample protocols, and, from this, the development, evaluation and proposal of a method for retrospective reporting of severity.

Each of these is considered in turn below. In addition, as part of the process in (ii), a number of illustrative examples of protocols have been collected and 'coded' for severity, and these are presented in Appendix 2.

4 Questionnaire findings

The questionnaire was distributed to all active project licence holders in nine designated establishments. A total of 460 licence holders were approached, and 168 completed questionnaires were received, representing 171 project licences (37% return). A detailed analysis of the results is presented in Appendix 1, and key points are summarised below. The findings helped to inform the Working Group's deliberations, and are drawn on where appropriate in the following discussions.

4.1 Current practice in recording adverse effects

- 4.1.1 The majority (79%) of project licence holders who responded to the questionnaire report that they "always" or "sometimes" make paper or computer records of the adverse effects actually experienced by the animals in the protocols they use. This rises to 88% of those whose project licences involve moderate and/or substantial protocols.
- 4.1.2 Some of those who report that they "never" record actual adverse effects in moderate or substantial protocols point out that, in practice, their protocol(s) cause significant adverse effects only very rarely (a moderate limit being regarded as a "safety net") or that the adverse effects of the protocols are very consistent/standardised.
- 4.1.3 Those who "sometimes" record actual adverse effects suggest that they do so when necessary for experimental or animal welfare reasons, including when:
- the adverse effects look likely to approach the severity limit;
 - the adverse effects are uncertain or difficult to predict;
 - a protocol causes unanticipated or unusual adverse effects in practice;
 - a protocol involves novel techniques, or techniques that are new to the project team or the establishment;
 - the adverse effects are part of the desired effect of the procedure;
 - characterising the phenotype of new mutant or transgenic animals.

4.2 Ease of Returning information currently collected by severity category

- 4.2.1 Around two in five (39%) of the respondents who currently record actual adverse effects say that they do this in a way that would allow them easily to Return the numbers of animals used by category of severity of actually experienced (mild, moderate, substantial).
- 4.2.2 The difficulties cited relate mainly to the extra *burden* this would pose, particularly because the information is not necessarily recorded in a *format* that would allow easy extraction of data that could enable the severity to be categorised. Some project licence holders also note that the information might be held in a number of different *locations*, recorded by different people for different purposes. Respondents also point to a number of areas of uncertainty, such as dealing with protocols involving breeding GA mice, and categorising adverse effects not directly connected to the protocol, including unexpected deaths of animals, or deaths due to ageing (see 7.2.3 below for further examples and discussion).

4.3 Ease of recording adverse effects by severity category

- 4.3.1 Respondents were asked to describe how they record (or would record) the severity of adverse effects experienced by the animals (as mild, moderate, substantial) and, as a follow-up, whether they foresee any difficulties with the method described. None of the respondents says that they already specifically and routinely record severity *by category* – though the use of clinical scoring schemes for implementation of humane end-points could result in such records.
- 4.3.2 Nevertheless 62% of respondents see no particular difficulties in making such records. Of the 38% who foresee difficulties, the main source of concern, again, is the extra burden this would bring and, in some cases, how this would detract from their obligations to science and animal welfare. Particular concerns are also expressed about the subjectivity of such assessments, as well as perceived

difficulties in dealing with GA animals and non-mammalian species (e.g. zebra fish), and implications for the complexity of the Returns form.

4.4 General concern

4.4.1 On reading all the replies to the questionnaire, an impression emerges that there is widespread anxiety that retrospective reporting of severity would mean that adverse effects (or lack of them) on *every* animal in *every* experiment/procedure would *necessarily* have to be recorded and categorised, with all the time and paperwork that would entail, and that this would have no discernible scientific or welfare benefit.

4.4.2 If, however, the severity of adverse effects can be predicted in advance of a study, it is possible that in practice recording could be much simplified. This is straightforward for protocols with mild severity limits, in which all animals should experience, at worst, mild effects. In moderate and substantial severity limit protocols, and depending on the nature of the work involved, it may also be possible to predict in advance of a study the category of severity of adverse effects that a given animal (or group of animals) will experience. For example, in a 'moderate' protocol involving five treatment groups, it might be possible to predict that four treatment groups will experience only mild effects, but that one group is likely to experience moderate effects.

4.5 Capacity to predict severity in advance of a study, then report 'exceptions'

4.5.1 Responses to the questionnaire suggest that such predictions frequently are possible – bringing the possibility that the process of recording actual severity could in these cases be reduced to recording only those animals that deviate from such expectations.

4.5.2 Overall, around three quarters (74%) of the 132 respondents whose project licences involve moderate and/or substantial severity protocols say that they can "usually" predict the severity of adverse effects in the procedures involved; and nearly 1 in 5 (19%) say that they can do this "occasionally". However, 7% say that they can "hardly ever" predict in this way. Examples of the latter two categories are listed in the detailed analysis in Appendix 1 (attached separately) Table 8b.

4.6 Views on subdivision of the 'moderate' severity category

4.6.1 Overall, nearly three-quarters (74%) of the 154 project licence holders who gave their views on this matter feel that the category should not be sub-divided. 24% think that subdivision would be helpful, and 2% are uncertain, suggesting that sub-division could perhaps be helpful.

4.6.2 The answers here are particularly interesting for what they reveal about the uncertainties and difficulties that respondents have in applying the current (or more probably, *any*) system of severity classification. These relate mainly to the 'subjectivity' of the existing classification and the need for better guidelines and examples to illustrate how to assign the categories. Those who oppose sub-division tend to argue that increasing the number of categories would over-complicate an already difficult-to-apply system (particularly at the boundaries between categories, where, it is felt, very difficult judgements have to be made) and/or would not benefit animal welfare.

4.7 Ways of improving the annual Returns procedure

Mindful that adding a requirement to the annual Returns to report severity retrospectively would add another layer of complexity to an already convoluted, and,

for some, difficult and time-consuming process, licence holders were also asked whether there are any ways in which the codes or other aspects of the procedure could be modified to make it easier for them to report animal use under their licence(s). 39 of the 101 licence holders who responded to the invitation to comment on this aspect suggested changes to the Returns process. These include:

- making the codes more intuitive (e.g. words rather than numbers), and providing more examples and better guidance in this and other areas (e.g. re-use);
- allowing (easier) submission on-line, collecting *only* data that is 'meaningful' in terms of animal welfare;
- simplifying and clarifying the reporting of use of GA animals;
- customising the Returns process to fit particular areas of work more easily; and
- re-arranging the hierarchy of the Returns form so as to avoid having to answer 'no' repeatedly.

A further 11 respondents appear to feel dissatisfied with the present process, but are uncertain what if anything might be changed – primarily because they find it difficult to envisage how the current complex procedure might be simplified. 51 respondents say that they have no particular problem with the Returns, and the remaining 67 licence holders who filled in the questionnaire did not comment on this final question.

5 Designing a system for retrospective reporting of severity

5.1 Assessing and recording *cf.* reporting adverse effects and their severity

5.1.1 Recognising, monitoring and taking steps to avoid or alleviate adverse effects experienced by laboratory animals is always vital, whether or not the effects (or absence of them) are recorded, categorised and/or reported.

5.1.2 As responses to the Working Group's questionnaire emphasise, current practice in recording actual adverse effects varies according to context: primarily according to the severity of the effects concerned and their predictability, as well as the type of work involved (see 4.1.2 above, and Appendix 1, Section 2.2). If a system of retrospective reporting of severity is brought in, it will be possible for there to continue to be variation in practice in the recording of actual adverse effects; but there will need to be a *common* system for *reporting* (and Returning) the severity of the effects for collation by the Home Office and publication.

5.1.3 In this section, we consider the possibilities for developing a *reporting* system capable of providing meaningful statistical information about actual severity experienced by animals in the full range of experimental contexts. We consider further the possible impact of such a system on *recording* practice (including associated bureaucracy) in Section 9 below.

5.2 Characteristics of a system for retrospective reporting of severity

Early in its discussions (and later in light of findings from the questionnaire), the Working Group identified a number of criteria which any system of retrospective reporting would have to meet. That is, it should:

- (i) be capable of generating data that provide as accurate and complete a reflection of the severity of actual adverse effects as possible – so as enable presentation of meaningful information in the public domain, and, importantly, neither under- nor over-estimate the severity of the effects concerned;
- (ii) following from (i), be capable of capturing both the intensity and duration of the adverse effects;

- (iii) be easily understood and straightforward to apply in practice, including being:
 - able to deal with the full range of contexts in which laboratory animals are used, and the full range of adverse effect profiles encountered; and
 - supported by clear guidelines, including worked examples;
- (iv) be minimally bureaucratic and maximally expedient (that is, provide maximum information for least effort), and/or have its extra burdens 'mitigated' by reductions in reporting requirements elsewhere in the Returns procedure (see Section 9 for further discussion);
- (v) bring animal welfare and/or scientific benefits, in addition to gains in terms of openness and public accountability (see Section 10 below);
- (vi) in relation to (v), generate information which has value for on-going, interim and retrospective ethical review by local Ethical Review Processes.¹⁰

6 Evaluating possible models for retrospective reporting

A number of possible models for retrospective reporting were identified and, in light of consideration of several detailed case studies, evaluated against the criteria listed in 5.2 above. The findings and conclusions are reported below.

6.1 Reporting number of animals used by protocol severity limit

6.1.1 Such a method, which has been suggested and debated in several different forums,¹¹ would be 'retrospective' in the sense that the use of animals is reported after the event. The system would offer some advantages, in that:

- the data so generated and reported would be an improvement on the publication of severity bands, providing information at the protocol rather than the overall project level, and therefore coming a little closer to reflecting the likely experience of individual animals;
- the reporting process would also bring minimal extra burden, simply requiring reporting of numbers of animals against criteria which, like other codes in the annual Returns, are fixed when the procedure is started – so capable of being dealt with just like any other of the 'rows' in the Returns form.

6.1.2 Set against these limited benefits, however, is a 'fatal flaw', in that these data would not reflect actual adverse effects; indeed, would greatly over-estimate actual severity¹². The severity limit of a protocol is just that: it sets an upper limit to the suffering that it is permissible to cause to an animal in that protocol. The limit therefore represents a 'worst-case scenario', which many (and in some protocols, most or all) animals would not be expected to experience. As a result, this method of reporting would provide seriously misleading information.

6.1.3 Conclusion:

Unfortunately, this method will not provide the 'quick fix' that might at first sight seem possible. If the extent and degree of laboratory animal suffering is to be represented

¹⁰ See LASA discussion paper: Jennings, M. and Howard, B. (eds) (2005). *Guidance notes on retrospective review: A discussion document prepared by the LASA Ethics and Training Group*. LASA: Tamworth www.lasa.co.uk/position_papers/publications.asp

¹¹ See, for example, Boyd Group/RSPCA (2004). *Categorising the severity of scientific procedures on animals: summary and reports from three round-table discussions*. RSPCA: Horsham <http://www.boyd-group.demon.co.uk>; also *Analysis of responses to the APC Consultation Paper on Statistics of Scientific Procedures in Great Britain* (2005) – in press.

¹² The above documents conclude similarly.

in the public domain, there is no alternative but to report, by some means, the severity of the *actual* adverse effects caused to the animals.

6.2 Reporting using a single code for each animal to indicate maximum actual severity

6.2.1 At its simplest, in the UK such a system would involve reporting the maximum severity of adverse effects actually caused to each animal subjected to a procedure, using the codes 'mild', 'moderate', 'substantial' or 'unclassified', as defined in Home Office guidance on the application of severity limits¹³.

6.2.2 This method of reporting was tested against the first example of 'real' severity data brought to the Working Party. The case is described in Box 3 below.

Box 3: Case Study 1 – Efficacy of oncology compounds

The work involved assessing whether novel compounds were effective in preventing tumours growing in mice. Both raw data and retrospective severity assessments for two studies involving different tumour lines were presented.

Study design	Inject tumour cells into mice; measure initial tumour growth; randomise animals to control and experimental groups. Then, daily dosing of experimental groups with a novel anti-cancer compound, and sham-dosing of control groups for 14 days (Study 1) or 21 days (Study 2). A reduction in tumour size in experimental animals compared to control animals indicates the anti-cancer compound is effective.
N animals	90 mice in Study 1; 95 mice in Study 2.
Protocol severity limit	'Moderate', primarily because the compounds could have unexpected adverse effects
Monitoring	The main indicators of condition are tumour size, body weight and clinical observations. Mice are observed at least twice daily throughout. They are weighed daily for the first week, thereafter they are weighed when tumours are measured (either every other day, twice weekly or three times a week). If at any time mice lose more than 8% initial body weight, they are weighed daily. Body weights and tumour sizes are recorded in an on-line database, which includes a field for commenting on any other clinical signs observed at daily checks.
Assessing severity of adverse effects	There are ERP-approved site guidelines for translating weight-loss effects into severity categories. As a rule-of-thumb, these are: Mild <10% weight loss; Moderate 10-20%; and Substantial >20% weight loss.
Predicted severity in practice	The vast majority of mice are expected to suffer 'mild' effects in these studies, because steps are taken to mitigate potential adverse effects, including implementation of humane end-points – e.g. a limit is set on tumour size and condition, beyond which, animals are humanely killed.
Actual severity observed	In Study 1, the condition of all 90 mice remained within the 'mild' severity category. In Study 2, the condition of 91 mice remained within the 'mild' severity category; 3 mice suffered 'moderate' adverse effects (>10% weight loss in one; and ulceration of the tumours in two animals, which were humanely killed); and one mouse suffered 'substantial' weight loss and was humanely killed. In all, then, 98% of mice were caused mild effects.

6.2.3 As Box 3 suggests (and setting aside resource implications for the moment), reporting maximum severity using the 'standard' categories, mild, moderate and substantial proved relatively trouble-free for Case Study 1, resulting in data that provided a meaningful assessment of the adverse effects caused (when supported by a description of how severity is related to the different degrees of weight loss).

¹³ Home Office (2000). *Guidance on the operation of the Animals (Scientific Procedures) Act 1986*. HC 321. TSO: London. www.official-documents.co.uk/document/hoc/321/321.htm

6.2.4 Case Study 1, however, is a comparatively 'straightforward' example, because:

- the severity effects are easily measured/assessed, relatively consistent between and within animals (varying little over the course of the procedure) and generally predictable (this means that, for the last two reasons, the work also lends itself well to 'exception' reporting); and
- the severity data are part of the experimental outcomes (part of the purpose of the science), and so the investigators are motivated to assess and record the information for scientific, as well as animal welfare and 'compliance', purposes.

6.2.5 As further examples were brought to the Working Group, and, in particular Case Studies 2 and 3 described in Boxes 4 and 5 overleaf, it became apparent that indicating maximum severity using the simple lexicon 'mild', 'moderate' and 'substantial', (as currently defined) would not adequately capture the effects of more complex, longer-term, procedures that generate more variable severity profiles over time.

6.2.6 Consideration of these more complex examples highlighted two general difficulties with the system of reporting described above:

- (i) The terms 'mild', 'moderate' and 'substantial', as currently defined, do not explicitly embrace duration (only intensity) of adverse effects¹⁴. As Cases 2 and 3 illustrate, the reporting system must take into account both aspects of severity in order to provide a meaningful assessment of actual adverse effects (see also 5.2 above). Thus, in order to accommodate more complex studies, and for comparison between different kinds of work, guidance would be needed on how to balance these two aspects of suffering in ascribing the codes retrospectively. An example of how this might be done is provided by the system for retrospective reporting of severity in Switzerland, summarised in Appendix 3¹⁵.

However, even with such modification, using just three codes (plus unclassified) to encompass both intensity and duration of adverse effects limits descriptive power and brings concomitant difficulties in interpretation of data reported retrospectively. The problem lies in the combining of more substantial but relatively short-term effects with milder but longer-term effects into single codes, as happens in the Swiss severity categories 2 and 3. Can these different effects really be considered 'equivalent' kinds of suffering, as the Swiss coding would suggest? By this system, Cases 2 and 3 above would both have been classified as causing Degree of Severity 3. Had the insult in Case Study 2 been somewhat less intense, that protocol would have been classified as Severity 2. Neither coding seems adequately to capture the difference between the two scenarios.

- (ii) By the same token, for longer-term procedures involving more complex interventions (such as in Cases 2 and 3 above), reporting only the *maximum* severity of adverse effect caused to an animal ignores other parts of the procedure which, depending on circumstances, can have greater influence on the overall experience of the animal. To capture the latter, some kind of estimate of total suffering is needed – perhaps what might be thought of as the 'area under the curve' in a graph of severity of adverse effect over the time course of a procedure.

¹⁴ Setting aside any concerns about nomenclature, it is clear that use of the code 'unclassified' would be relatively unproblematic (this code being reserved for animals used in protocols performed solely under general anaesthesia, from which the animals are not allowed to recover).

¹⁵ Degrees of severity 1, 2 and 3 could be considered equivalent to the UK categories 'mild', 'moderate' and 'substantial'. In the UK, procedures having the degree of severity 0 would not be regulated, and therefore would not need to be Returned.

Box 4: Case Study 2 – a protocol involving non-human primates

Study Design:	A non-human primate is surgically prepared with an experimental device and then injected with a compound to induce clinical signs. Clinical signs are ameliorated or abolished by periodic dosing of therapeutic agents or by use of the experimental device. Experiment continues for up to 18 months.
N animals:	1 non-human primate per year.
Protocol Severity Limit:	Substantial, due the clinical signs that may be induced initially before the treatment regime begins, or if the treatment is ineffective.
Monitoring:	A sophisticated and comprehensive scoring system is used to accurately assess the intensity of clinical signs and to titrate the treatment doses to those most appropriate for the study.
Assessing severity of adverse effects:	Numerical scores from the scoring system equate to levels of severity on a basis decided in consultation with the ERP.
Predicted severity in practice:	The nature of the study is intended to prove and optimise the treatment regime, so after the initial sequence of the experiment the animal may well experience no detectable adverse effects for much of the time.
Actual severity observed:	An initial 'peak' of substantial severity was observed in the initial sequence of the experiment, as a result of the induction of clinical signs. Once on to the treatment sequence and for the remainder of the experiment, severity was never higher than moderate and was usually no higher than mild.

Box 5: Case Study 3 – a protocol involving ferrets

Study Design:	<p>Ferrets are trained to perform behavioural tasks in return for fluid reward, necessitating controlled access to water and sometimes single housing.</p> <p>Under general anaesthesia and with perioperative analgesia some ferrets are then prepared by implantation of experimental devices, with either reversible or irreversible deactivation of areas of the brain.</p> <p>Behavioural testing continues to investigate the effects of the interventions. The experiment continues for up to 12 months.</p>
N animals:	Approximately 10 per year.
Protocol Severity Limit:	Moderate, reflecting surgical interventions with appropriate anaesthesia and analgesia, and extended periods of controlled access to water.
Monitoring:	Specialised monitoring schemes for assessing post-operative recovery from surgical procedures. Clinical assessment and body weight records to monitor general health and body condition during periods of restricted water, as well as behavioural assessment.
Predicted severity in practice:	Anaesthetic and analgesic strategies keep surgical procedures within the moderate category. Extended periods without free access to water and single housing reflect continuing severity likely to vary between mild and moderate.
Actual severity observed:	Several different experiments are carried out under the same protocol, some of which are mild throughout, others are moderate throughout, and yet others vary and fluctuate between the two.

6.2.7 Conclusion:

The Working Group is agreed that, whilst the kind of a single-code scheme explored above has some merit, particularly in its relative simplicity, it cannot deal adequately with relatively long-term procedures that have complex severity profiles. In particular, the system needs modification in order to capture duration as well as intensity of adverse effect.

Combining intensity and duration as in the Swiss system is likely to cause significant difficulties in interpreting the data reported. A possible solution would be to increase the number of categories, so as to avoid having to combine into a single code more substantial but relatively short-term effects and milder but longer-term effects.

In order to provide as meaningful an assessment of severity as possible, there is a need to capture some measure of the *overall* experience of the animal over the duration of the procedure.

6.3 Reporting using an intensity-duration severity grid, to indicate either maximum or overall actual severity

6.3.1 In light of the preceding discussion, the Working Group proposed a simple grid system, which would give each possible combination of intensity vs. duration of adverse effect a unique severity code. This is shown in Figure 1 overleaf.

Figure 1: Grid system for reporting intensity and duration of adverse effects

Intensity of adverse effects	Duration of adverse effects		
	Short	Medium	Long
Mild	A	B	C
Moderate	D	E	F
Substantial	G	H	I

6.3.2 The grid classifies intensity according to the current labels, mild, moderate and substantial, and suggests that duration of severity also be divided into three categories. For simplicity, the latter are labelled short, medium and long – but would need further definition in practice. Letters rather than numbers are used for each of the resulting nine possible intensity-duration combinations in recognition of the qualitative judgements involved, and in order not to imply that, for example, Moderate-Short adverse effects (which would be coded 4 in a numerical scheme) impose 'twice' the severity of Mild-Medium duration adverse effects (which would be coded 2).

6.3.3 The grid overcomes the difficulties discussed in 6.2.6(i), by explicitly including both intensity and duration of effect in the classification of severity, and by avoiding combining different intensity-duration combinations within a single code (as the Swiss system does). It works well for the kind of straightforward example explored in Case Study 1.

6.3.4 On its own, however, it is less successful in dealing with the problem stated in 6.2.6(ii) above, and illustrated by Case Studies 2 and 3:

- if the grid is used to report *maximum* intensity of severity and its duration within a protocol, the problem of capturing important differences in *overall* severity (particularly in longer-term procedures), described above, will remain;
- alternatively, using the grid to report *overall* severity, over the whole duration of the procedure, gives rise to another kind of difficulty. This would require an estimate of the 'average' intensity over the whole procedure, which would mean, for example, that 'mild' effects in one part of the procedure would be considered to mitigate 'substantial' effects in another part. As a result, this approach would tend to downplay the peak suffering actually caused.

6.3.4 Conclusion:

A single intensity-duration severity grid works well with procedures that have relatively simple severity profiles, but is less successful in providing a meaningful reflection of severity in more complex, longer-term procedures.

6.4 Reporting using two intensity-duration severity grids, to indicate maximum severity and severity over the remainder of the procedure

6.4.1 In order to capture severity profiles over the whole duration of procedures more effectively, the Working Group then explored the possibility of using two intensity-duration severity grids for reporting: one to report *maximum* severity and its duration within a procedure, the other to report the '*remaining*' severity, i.e. severity over the rest of the procedure, disregarding the 'peak'. The scheme is shown in Figure 2 overleaf.

6.4.2 It was initially suggested that the second grid be used to report severity over the duration of the *whole* procedure (i.e. should include the maximum severity and all other adverse effects). However, when tested against real scenarios, the information generated did not give as clear a picture as using the second grid to report '*remaining*' severity (i.e. severity over the whole procedure *minus* the maximum severity), and was not as intuitive to researchers who attempted to use the system.

6.4.3 The grids shown in Figure 2 now include suggested definitions of 'short', 'medium' and 'long' in each case. The Working Group suggests that the first grid should be taken to refer to the *total* duration of the maximum effects within the procedure. So, if a procedure (at its maximum effect) includes several episodes of moderate intensity adverse effects each lasting less than a day, but having a combined duration of more than one day, these effects would be reported under code E in Grid 1 (see also 7.2.3iv).

6.4.4 Using this system, the examples already considered in detail by the Working Group would be categorised as follows (for *most* animals):

	Grid 1	Grid 2
Case Study 1:	C	B
Case Study 2:	H	C
Case Study 3:	E	F

6.4.5 It is important to note that, although at first sight the two grid scheme appears to yield 81 code combinations, far fewer are possible in practice, because the '*remaining*' severity cannot be greater than the maximum (i.e. the intensity of adverse effects in Grid 2 cannot be greater than those in Grid 1; furthermore, if intensity is the same throughout the procedure, then the duration in Grid 2 cannot be greater than in Grid 1).

Figure 2: Two-grid system for reporting the severity profile of a procedure

Grid 1: *Maximum* severity within the procedure

Maximum intensity of adverse effects	Duration of maximum intensity of adverse effects within the procedure		
	Short ≤ one day	Medium >1 and ≤ 7 days	Long > 7 days
Mild	A	B	C
Moderate	D	E	F
Substantial	G	H	I

Grid 2: *Remaining* severity, over the rest of the procedure (*less* the maximum)

Overall intensity of adverse effects in rest of procedure	Duration of the remainder of the procedure		
	Short ≤ 7 days	Medium >7 and ≤ 28 days	Long 28 days +
Mild	A	B	C
Moderate	D	E	F
Substantial	G	H	I

6.4.7 Considering possible code combinations in more detail, two particular questions might be raised:

- (i) if the suggested durations shown in Figure 2 are adopted, how should procedures lasting less than or equal to a day be dealt with? Since there would be no need to record 'remaining' severity (Grid 2) in such circumstances, one possibility would be to record an X for Grid 2 for these short-duration procedures (e.g. DX);
- (ii) How should procedures that have the same severity throughout their duration be dealt with? Would a single code be better here? The Working Group's initial view is that, since the codes generated by application of the two-grid scheme would be an accurate description of the severity profile of the procedures involved, the classification system should not be complicated by addition of other special codes, which would add further columns in the Returns and make the grid-scheme more difficult to apply in practice. Rather, the implications of these particular codes (e.g., with the durations shown in Figure 2: AA, BA, BB, CB, CC, DD, ED etc) for severity in practice should be explained in the text accompanying presentation of the data in the *Statistics*.

- 6.4.8 Early, albeit limited, feedback from the licence holders in the Case Studies already described suggests that the two-grid system is understandable, intuitive to apply, and acceptable in terms of its capacity to portray the severity of adverse effects experienced by animals in more complex procedures.
- 6.4.9 This impression is supported by the Working Group's own testing of the two-grid system over a range of different kinds of procedure. Each of the examples considered by the Working Group is described briefly in Appendix 2, alongside a note of the coding decided upon and any difficulties encountered in assigning the code.
- 6.4.10 Note that applying the two-grid system in practice requires only four independent decisions, two of which are subjective judgements (the peak and remaining intensity of adverse effect), and two of which (the durations) are more objective. Once these decisions have been made, the required coding can simply be 'read off' the grids, mechanically, without the need to exercise further judgement.
- 6.4.11 Conclusion:

In light of the foregoing discussions and testing against the examples shown in Appendix 2, the Working Group concludes that the two-grid system can potentially provide a representative picture of intensity and duration of severity over a wide range of different procedures.

Although at first sight the two-grid system appears to bring a large number of candidate codes for a given procedure, the possibilities will be fewer in practice – being limited both by the nature of the coding scheme itself and by the design of the protocols involved. For many studies, it is likely that appropriate codes will be predictable in advance, or at least narrowed down to a small range of choices. This was the experience of the licence holders in Case Studies 1-3. Once it was realised that, for these protocols, reporting could be approached by: (i) predicting severity codes and then (ii) identifying and coding 'exceptions', the task became more manageable. Nevertheless, it is clear that adoption of any such system of retrospective reporting will have significant resource implications. See Section 9 for further discussion.

6.5 Further development of reporting system

Developing a system for retrospective reporting of severity involves balancing (potentially) competing requirements: (i) to enable as accurate and meaningful a coded description of the actual severity of adverse effects as possible, applicable to a wide range of different protocols; and (ii) to minimise the bureaucratic burdens imposed by this additional reporting requirement. Striking the 'right' balance is difficult.

Further consideration should be given to the possibility of combining some codes generated in the two-grid scheme, so as to simplify the process of reporting and provide easier 'de-coding' of the data when they are presented in the public domain. The proposed system will also require further evaluation, in order to assess its capacity to deal with a wider range of protocols, to gain more insight into the likely impacts on project licence holders and their establishments, and, if necessary, to modify the proposals in light of the feedback obtained. See also Section 11 below.

7 Issues for further consideration/resolution

In considering how the two-grid coding scheme might be applied in practice, the Working Group has identified a number of points for further consideration. [Note that resource impacts are considered in a separate section (9) below.]

7.1 Should there be a 'no effect' category?

- 7.1.1 This question was raised in responses to the questionnaire and also via the Working Group's evaluation of examples. It was suggested that there should be a special category (an extra 'intensity' code) to reflect occasions when, in practice, no overt adverse effects are caused to the animals, e.g. in control or 'no effect' test groups, or when animals are 'simply' maintained in standard housing and husbandry conditions for observation following an intervention¹⁶. However, against this, it might be objected that there is always a measure of adverse effect caused purely as a result of animals being maintained in the laboratory environment, being handled and so on; and/or that a precautionary approach should be adopted, in that the absence of 'overt' signs should not be taken to mean that there are necessarily no adverse effects.
- 7.1.2 The Working Group suggests that labelling the 'mild' intensity category as 'up to mild' would accommodate both of these perspectives within the two-grid coding scheme. See also comments in Section 7.2.3(v) below, in relation to what might be a 'special case' in this context.

7.2 Need for further guidance and examples on how to assign codes

- 7.2.1 It is clear that, for public information purposes, retrospective severity coding would need to be applied as consistently and reliably as possible. Yet, responses to the Working Group's questionnaire suggest that project licence holders feel that there are already considerable uncertainties in applying the current severity classification system (e.g. when deciding severity limits and bands for project licences) and that judgements at the boundaries of the categories can be very difficult. Comments made in responses also suggest that licence holders would feel considerable anxiety about 'getting it right' when reporting severity retrospectively¹⁷.
- 7.2.2 Any scheme for retrospective reporting will therefore need to be supported by detailed guidance, including a catalogue of worked examples, covering the full range of species and a wide variety of regulated procedures. Since the recording of intensity of severity within each grid for a particular study type or model may vary between individuals and establishments, a glossary of severity codes for a range of protocols and outcomes, agreed by project licence holders, is essential. Such guidance, examples and glossary would also be valuable in explaining the data when they are presented in the public domain.
- 7.2.3 The guidance will also need to include agreed 'rules' on how to code some contentious or uncertain effects, including the following:
- (i) *How to code non-procedure related death and sudden death without prior suffering*
There is no doubt that a procedure which causes suffering resulting in death of an animal should be classified as substantial severity. But there is considerable uncertainty about how to view the severity of non-procedure related deaths, such as those which occur in aging colonies of animals; and sudden death without prior suffering, such as occasionally occurs in a genetically modified strain of mice with enlarged hearts, and another in which individuals occasionally suffer an aortic

¹⁶ See Example 8 in Appendix 2 for illustration

¹⁷ Participants in a recent focus group made similar observations – see Boyd Group/RSPCA (2004) *op cit*.

aneurism causing sudden death. At present, the first of these sudden death cases has a severity limit 'moderate', whilst the second has a severity limit 'substantial'.

Moreover, sudden death can also occur as a direct result of an intervention during a procedure (e.g. due to the effects of a compound in a toxicity study), but, because of the short time-scale involved, without detection of any preceding clinical signs. This was the case in a fourth example considered in detail by the Working Group, which is outlined in Box 7 overleaf (Case Study 4).

Basing application of the retrospective reporting system solely on clinical signs *actually observed* would result in the coding (GA), as shown in the Box, for small numbers of animals 'found dead' in the highest dose group of an acute toxicity study, without previous clinical signs being observed (despite regular observation). Some members of the Working Group, however, would prefer to code this situation as GG. The latter coding would assume that there was short-term substantial suffering prior to death that was not evident from the last set of clinical observations, and so would give the benefit of the doubt to the animal. Those who argue for a GG coding suggest that a GA coding in such circumstances (based solely on signs actually observed) would be unacceptable to a lay-person, given that the objective of this study type is to cause major adverse effects. They ask what would be the downside of giving the animal the benefit of the doubt in examples such as this?

Clearly, for consistency in retrospective reporting, there will need to be agreement on how death of animals should be coded in all the different circumstances outlined above.

(ii) *How to code the severity of surgical procedures*

At present, the majority of surgical procedures are conducted under 'moderate' protocol severity limits, and there appears to be a divergence of opinion on whether the actual severity of such procedures should always be coded as 'moderate' intensity, or whether, with good analgesia, a 'mild' code could sometimes be appropriate. For example, in the case of scrotal sac vasectomy in mice¹⁸, some would argue that, with an effective analgesic regime, only mild effects will be caused in practice; others, however, suggest that, since it is impossible to be sure of the efficacy of the analgesia, the benefit of the doubt should lie with the animal and the intensity of severity of the procedure coded as moderate. Guidance on retrospective assessment of severity will need to adopt a clear position in this and similar cases.

(iii) *Factors to take into account in Grid 2*

Should Grid 2 take into account contingent effects, such as those pertaining to source of the animals, their transport to the laboratory, quality of 'standard' housing and husbandry – or just effects directly caused by the procedure itself? The Working Group's initial view on this question is that the effects taken into account in the severity coding should be limited to those caused by regulated interventions, plus only those contingent effects which are under the control of the licence holders who direct and/or carry out the regulated work concerned.

(iv) *How to deal with cumulative effects*

As noted, the Working Group proposes that Grid 1 should refer to the *total* duration of the maximum effects caused to the animals (see 6.4.3 for an example), even if they are spread over several different episodes. This raises the question of how repeated effects can influence intensity of severity. Might intensity of severity be 'cumulative' in some procedures (e.g. repeated injections), in that several episodes of, say, 'mild'

¹⁸ The procedure is described in Example 2 in Appendix 2.

Box 7: Case Study 4: Acute toxicity study

In pharmaceutical drug development acute toxicity studies in rodents are a regulatory requirement. The objective is to identify a dose causing major adverse effects after a single dose up to a limit of 2000 mg/kg. The animals receive a single dose of the test compound followed by up to 14 days observation post-dose.

The following two examples have been selected at random.

Study Design	A dose-range (up to 2000 mg/kg) is initially investigated in 1 male and 1 female animal <i>per</i> dose level. A group of 5 + 5 animals is dosed at the anticipated maximum non-lethal dose. Occasionally it may be necessary to add an additional group of up to 5 + 5 animals. Typical study sizes range from 10 to 25 animals.
N animals	Study 1: 22 rats. Study 2: 10 rats
Protocol severity limit	Substantial, primarily because these will be the highest doses investigated in any toxicity study and major adverse effects are anticipated in a small number of animals
Monitoring	The animals are observed regularly every day (particularly on the day of dosing) for any signs that indicate a deviation from normal appearance or behaviour. In addition, body weight is recorded regularly.
Assessing severity of adverse effects	There are ERP approved site guidelines for translating clinical observations and weight-loss effects into severity category
Predicted severity in practice	The vast majority of animals are expected to experience moderate effects in these studies, because steps are taken to mitigate potential adverse effects usually by implementation of humane endpoints. A smaller number of animals will experience substantial effects, as the purpose of the study is to identify major adverse effects up to a limit of 2000 mg/kg. However, the number of animals experiencing substantial effects is kept to a minimum because the study design involves the dose range being investigated in one male and one female. In studies where compounds of anticipated low toxicity are used and where the limit dose only is given (2000 mg/kg) then the majority of the animals may only experience mild effects.
Actual severity	<p>In study 1 six animals had a maximum intensity of adverse effects of mild (B) and eleven of moderate (E). Five animals had a maximum intensity of adverse effects of substantial (G) over a time period of less than 24h. The effects and coding based on clinical signs <i>actually observed</i> are described in the Table below. <i>See main text for an explanation and discussion of the GA coding for the two animals found dead without previous clinical signs observed.</i></p> <p>In study 2 all animals had a maximum intensity of adverse effects of mild (A) and appeared normal throughout the study. The severity coding was therefore AB.</p>

Adverse effects in Study 1	Severity coding (N animals)
Transient increased heart rate and slight pilo-erection. Observed on the day of dosing only. Animals appeared normal for remainder of the study.	BB (6)
Irregular heart rate, irregular respiration, pilo-erection. Seen intermittently over 24 h. Animals appeared normal for remainder of study.	EB (11)
Died on day 1. Previous observations included pilo-erection, decreased activity, pale, irregular respiration, eyes half shut, cold.	GG (1)
Found dead on day 2, no previous observations recorded the evening before.	GA (2)
Found dead on day 2. Previous observations included pilo-erection and cold extremities.	GD (2)

PPL's comments:

As this type of toxicology study is run on a substantial limit each animal should be assessed individually. Larger studies conducted on a moderate or mild limit may lend themselves to group assessment as long as clear guidelines and definitions were available. The value of the proposed system is that it reflects the actual suffering of each individual animal which is more accurate than using the current severity limit (substantial) and the data generated may be used to further refine study design and observation periods. It is clear from this small analysis that even with a study type that requires a substantial limit and that has a lot of public/scientific attention and criticism that there is a wide range of degrees of suffering with the majority of animals experiencing mild or moderate effects.

effects become 'moderate' by the last occasion; and, if so, how should these cumulative effects be coded?

(v) *How to report phenotypically normal GA animals bred but not otherwise used*
Breeding genetically altered animals (both GM and mutant animals) is regarded as a regulated procedure, whether or not the animals suffer adverse effects as a result of the genetic change, and whether or not they are subsequently used in another regulated scientific procedure. There are some particular uncertainties in considering how to report the 'severity' of breeding phenotypically 'normal' GA animals that are not otherwise used:

- what weight should contingent effects, such as those outlined above, play in the severity assessments, particularly where phenotypically 'normal' GA animals are concerned?
- should there be a 'no effect' category for phenotypically 'normal' GA animals bred but not otherwise used?
- if not, and they are classed as 'up to mild', is it reasonable/meaningful to add a duration code for these animals, depending on at what stage in their life they are killed?¹⁹

(vi) *How to categorise 'harm' as compared with 'suffering'*
ASPA specifically regulates procedures causing 'lasting harm' to animals, as well as those causing suffering. It is possible to envisage 'harm' without suffering, as for example, in the case of transgenic mice which have smaller than normal external ears (pinnae). There is no evidence that the animals' hearing is affected, nor that the genetic alteration causes any other adverse effect. What weight should such 'harm in the absence of suffering' be accorded in retrospective reporting of severity: mild, moderate, or?

(vii) *Studies lasting more than a year and those crossing year-ends*
Clearly, actual severity has to be reported when a procedure is completed. The present Returns system, however, asks for records of procedures started in the given year, which means that severity data for procedures lasting more than one year and those crossing year-ends would be collected and reported in later year(s) than other data pertaining to those same procedures. Further thought therefore needs to be given to how best to integrate retrospective reporting of severity data with reporting of other data in the annual Returns. See Section 8 for options and further discussion.

7.3 Further developing consistency of approach in retrospective reporting

Clearly, however good the guidance, there is likely to be at least some variation in approach to retrospective reporting, and room for disagreement about how to assign the codes. Inevitably, an element of subjective judgement will be involved. In order to help in developing consistency of approach, it might therefore be helpful to subject such assessment to occasional 'audits'. These could take the form of workshops or other forms of discussion between licence holders in contentious areas, and could bring the added benefit of helping to promote refinement of the procedures involved.

8 Options for Returning data collected

Having proposed a method for capturing severity information retrospectively, the Working Group also considered how the information collected might best be Returned to the Home Office, along with other data that goes into the annual

¹⁹ See Example 5 in Appendix 2 for illustration.

compilation of *Statistics*. A number of options are possible, each having pros and cons.

8.1 Prospective Returns with additional severity data, or retrospective Returns incorporating severity data?

The current Returns (and resulting *Statistics*) count procedures when they are started (i.e. prospectively), whereas, obviously, data on actual severity are retrospective.

Option 1: Stick with the existing 'prospective' Returns system (suitably pared down – see 9 below), and ask project licence holders to report retrospective severity data separately from all other statistical information.

This option has the advantage that it would not require a major change in the current system of collecting data on animal procedures – being merely a 'bolt on'. The obvious disadvantage is that it would not be possible to cross reference the severity data to any other part of the published *Statistics*, and would lead to 'double counting' of procedures (at their conclusion for severity information, and when they are started for the rest of the Returns). There would therefore be considerable additional burdens on licence holders and establishments.

Option 2: Change the whole Returns and statistical publication so as to report all data retrospectively.

This would ensure comparability of data between the severity and other sections of the published *Statistics*, would avoid 'double counting' procedures, and would also enable information to be collected on the duration of procedures. Furthermore, retrospective reporting would count *in each year* procedures lasting more than a year and those crossing year ends, and so would enable better estimation of the numbers of animals actually *in use* in any given year (cf. procedures *started*, as the current system records)²⁰ – thus dealing with the difficulty noted in 7.2.3(vii) above.

The Animals (Scientific Procedures) Act 1986 merely requires that 'information' on animal procedures be published, so there is no regulatory impediment to a change in the method of reporting. However, such wholesale change would require project licence holders and their establishments to alter their systems for recording and Returning data, and, initially, would disrupt some of the historical trends currently monitored in the published *Statistics*.

As noted, if Option 2 is followed, the retrospective severity information could be collected as extra Row(s) in the Returns form, enabling data in the severity Row(s) to be cross-referenced to data in other Rows – e.g. primary purpose, legislative or other reason for the work, genetic status of the animal, use of anaesthesia, etc. The disadvantage would be that, with a relatively large number of severity codes (even in a one-grid system), the Returns process would become considerably more complex, because more columns would be generated, and therefore more onerous (see 9 below for further discussion).

8.2 Routine annual Return of severity information for all project licences, or selective/exemplar reporting?

With both Options 1 and 2 above, there is the possibility that, rather than requiring every project licence holder to Return severity data annually, sub-sample(s) of project licences in particular area(s) of work could be reported each year, and written up as exemplars in the *Statistics*. Although the severity information reported would not

²⁰ *Report of the APC Statistics Working Group – in press*. Note that the APC report also explores another (more effective) option for achieving the latter, but this is prospective, not retrospective.

capture all animal use in any given year, it could be presented in more 'depth' – so as to set the adverse effects caused to the animals in the context of the benefits sought from the work, for example²¹. However, this sampling method would probably not greatly reduce the effort required from project licence holders, since they would still have to *record* the information even they were not required to report it in a given year.

9 Resource implications of a retrospective severity reporting process

9.1 As might be expected, the Working Group's questionnaire to project licence holders reveals widespread concern about the time and other resources implications of an additional requirement to report severity retrospectively. Many already perceive the annual Returns process as overly complex and therefore difficult and onerous. Adding another reporting requirement (and, to boot, one that requires 'subjective' judgement to be exercised – with all the uncertainties and worries about 'error' that would bring) is likely to be perceived as the proverbial straw that breaks the camel's back.

9.2 Responses to the questionnaire also reveal that in many cases severity can be predicted in advance of studies, so that 'exception' recording methods are likely to be possible. Moreover, in circumstances in which severity is uncertain, the questionnaire responses confirm that severity data are in any case almost always recorded – either for experimental reasons (when part of the purpose of the science is to characterise or better understand the adverse effects and their cause) and/or for animal welfare reasons (to help in monitoring the animals and taking steps to minimise adverse effects observed) and/or for compliance reasons (to avoid inadvertent breaches of protocol severity limits).

9.3 Nevertheless, one should not under-estimate the time likely to be required to report actual severity carefully and thoughtfully. The limited data obtained in this pilot study confirm that meeting such a reporting requirement will be time-consuming, and therefore that the process will make significant resource demands on both the individuals and establishments concerned. For example:

(i) In Case Study 4, the researcher reports that the time taken to carry out individual assessment of severity is minimal for each study of this size: five minutes for the two studies. However, if large numbers of such studies are conducted over the course of a year, recording and reporting severity could bring considerable burdens.

(ii) A researcher in industry estimates that it will take around 20 minutes to record actual severity in a relatively straightforward toxicity study involving 80 to 100 animals. With upwards of 240 such studies conducted under a single licence per year, the total time required to record severity for just this one kind of toxicity test would be 80 hours (over 11 working days) annually²².

(iii) Another industry researcher reports that it took over an hour to assign codes for maximum severity in a study involving 164 mice to assess treatments of arthritic symptoms²³ (in which there is already intensive scoring of clinical signs); and that estimating remaining severity would take at least a further 30 minutes. About 10 such studies are performed per year, each involving

²¹ This was a point brought out in the licence holder and NVS/NACWO focus groups in the Boyd Group/RSPCA's recent study: participants argued strongly that information about the severity of procedures should be set alongside information about the purpose of those procedures, so as to provide an indication of the harm-benefit judgements involved.

²² See Example 8 in Appendix 2

²³ Example 7 in Appendix 2

approximately 150 mice. Therefore it is estimated that the extra regulatory burden over and above the current intensive scoring would be 15 hours per year to assess, plus 2 hours a year to Return the data to the Home Office. Training the licensee concerned was reasonably fast and so it is anticipated there will be little further reduction in the time required to assign scores in future studies.

- 9.4 Clearly, it would be valuable to obtain more information on resource impacts in practice, and to consider further how these might be mitigated (see Section 11). Early, albeit limited, feedback suggests positive reactions to the Working Group's proposed method of retrospective reporting; but also show that, without concomitant reduction and/or simplification of reporting requirements elsewhere in the annual Returns and the processes associated with the operation of the Act, that favourable reaction will evaporate very swiftly.
- 9.5 In this context, it is recommended that the principles outlined by the Government's Better Regulation Task Force are adopted, so as to maintain (or even reduce) the overall burden imposed on licence holders and others, such as animal care staff (a 'one-in, one-out' principle)²⁴. Possible opportunities to reduce and/or simplify processes associated with A(SP)A if retrospective severity reporting was introduced are suggested in Appendix 4.
- 9.6 In light of concerns about burdens on licence holders and establishments, further thought should also be given to the degree of 'accuracy' required in severity reporting. The Working Group notes that, in some other countries requiring reporting of adverse effects (e.g. Switzerland), data are reported in a more 'impressionistic' way²⁵.

10 Potential benefits of retrospective reporting of severity

- 10.1 In light of the efforts required to report severity data and for maximum 'buy-in' to any such system, it is important that the process should bring benefits to animal welfare and/or to science, in addition to gains in terms of openness and public accountability (see criteria listed in Section 5).
- 10.2 The Working Group's discussions, consideration of case studies and feedback from licence holders suggest that such retrospective reporting could bring the following additional benefits (some of which have already been alluded to in this report):
- (i) reporting actual severity in a systematic manner could help to promote on-going refinement of procedures, by:
- providing a focus for licence holders' own consideration of severity and possibilities for refinement;
 - facilitating proactive engagement of Named Persons with licence holders and others within the establishment on severity issues;

²⁴ Better Regulation Task Force (2005). *Regulation – less is more: reducing burdens, improving outcomes*, A BRTF report to the Prime Minister. The Cabinet Office, Publications and Publicity Team: London. Available via: www.brtf.gov.uk

²⁵ In this context, it is also interesting to consider the situation in The Netherlands, where information on the degree of discomfort experienced by laboratory animals also has to be reported after experiments have been performed. A summary of these data is published in Dutch annual statistics on animal experiments, according to whether the suffering experienced by the animals was: 'minor', 'minor/moderate', 'moderate', 'moderate/ severe', 'severe', or 'very severe'. Information on how the categories are applied is exchanged at meetings of Animal Welfare Officers and there is discussion aimed at promoting consensus (see FELASA Ethics Working Group report – *in press*; and to be made available via a link from www.felasa.org). See also: (VWA (2003) *Zo Doende 2003 Statistics of animal experimentation in the Netherlands*). VWA: The Hague. Available at: http://www.vwa.nl/download/rapporten/Dierproeven/zodoende_2003.pdf.)

- generating information that can be fed into on-going, interim and retrospective ethical review by local Ethical Review Processes;
- (ii) developing and refining the system of retrospective reporting should provide an opportunity to develop glossaries of clinical signs and their severities related to regularly-used procedures, which would help licence holders in monitoring animals and managing/reducing severity, as well as reporting severity information retrospectively;
- (iii) reporting severity data in the annual *Statistics* would help to ensure that accurate information on actual severity enters the public domain – at present there is little or no publicly available information to substantiate claims by the research community that many or most procedures are of 'mild' severity.

11 Suggestions for further work

11.1 This pilot feasibility study has been conducted over a very short time-scale and the above proposals have therefore been tested against a rather limited number of examples. If the Working Group's proposals are considered acceptable at this stage, they should then be piloted more systematically (e.g. 'across the board' in a small number of establishments representing a wide variety of different kinds of work).

11.2 An extension of this work is therefore proposed, in order to:

- consider simplification of the proposed reporting scheme (e.g. by combining categories to generate a smaller, more manageable and interpretable number of options for reporting);
- assess its appropriateness across a wider range of establishments;
- evaluate the consistency of response to the scoring system;
- identify a process for developing a glossary of severity codes for a range of protocols and outcomes (including examples of the use of farm animals and non-mammalian species);
- clarify a number of unresolved issues in the application of the coding scheme (see section 7 above);
- conduct a detailed impact assessment; and
- assess the value of the scheme in generating data that can assist public understanding of the severity of effects caused to animals used in regulated scientific procedures.

APPENDIX 1:

Summary of responses to the questionnaire: separate document

APPENDIX 2: EXAMPLES CONSIDERED BY THE WORKING GROUP

Case studies presented in the main report:

	Page
Case Study 1: Efficacy of oncology compounds	10
Case Study 2: A protocol involving non-human primates	12
Case Study 3: A protocol involving ferrets	12
Case Study 4: Acute toxicity study	19

Other examples considered by the Working Group:

1: Long term study of effects following multiple vaccinations

Marmosets were used to investigate potential long-term (18 month) effects following multiple vaccinations and other pre-treatments administered over one month. Parameters measured included cognitive behaviour, EEG, sleep, immunology, endocrinology and muscle strength. EEG (and thus sleep) were measured by means of pre-implanted radiotelemetry devices. Cognition and muscle strength were measured using home cage testing techniques whereby animals had a choice of whether or not to engage. Blood samples were taken on a monthly basis. Animals were trained to provide early morning urine samples when required.

Severity scores

Grid 1 – maximum severity – E – because of implantation surgery, anaesthesia and recovery

Grid 2 – remaining severity – C – because of monthly blood sampling over 18 months

The grid system would allow for individual outliers in such a study, e.g. those possibly showing adverse effects resulting from surgery or the blood sampling, to be recorded as appropriate.

2: Vasectomy of male mice for generation of pseudopregnant recipient mice

Mice are anaesthetised with isoflurane and a 5-7mm scrotal incision is made under aseptic surgical conditions. This allows access to the vas deferens via small incisions in the peritoneal layers, the vas is ligated with cautery then the incisions are closed at the peritoneal and skin levels. Mice are given a dose of buprenorphine for post-op analgesia. Recovery and subsequent days post-op are usually uneventful.

Mice are then retained for 12 months, caged individually but with females introduced overnight at least two days per week.

Severity scores

Grid 1 – maximum severity – A or D – depending on the view taken on the effectiveness of post-operative analgesia and giving the animals the benefit of the doubt – see Section 7.2.3(ii); definitely D or E if there are signs of infection and/or self induced trauma in the post-operative period

Grid 2 – remaining severity – C

3: Plasmid electrotransfer into skeletal muscle of mice

Procedure involves injection of a dilute solution of hyaluronidase (25 microlitres) into the anterior tibial muscle of a 12 week old male mouse. This is carried out under Hypnorm/Hypnovel anaesthesia. 2 hours later the mouse is re-anaesthetised with isoflurane, plasmid DNA is injected into the same muscle and calliper electrodes are applied to the skin and 10 twenty millisecond electrical pulses are given (175V/cm at 1Hz). Depending on the experiment the mouse may then be given an injection of buprenorphine as the analgesic.

Recovery is rapid with only transient (couple of hours) evidence of any lameness. There are no clinical signs of lameness 24 hours after the procedure.

Animals are generally killed 5 days later and the muscles removed for post-mortem analysis. In a few cases there is clear evidence of significant muscle damage on histological evaluation.

Severity scores

Grid 1 – maximum severity – A – if with analgesia and no damage; or possibly D/E if no analgesia and clear evidence of muscle damage on post-mortem

Grid 2 – remaining severity – A – if no major muscle damage; Or possibly D if clear evidence of major muscle damage on post-mortem.

4: Effect of pharmaceuticals that influence mood and behaviour on brain responses to environmental stimuli in mice and rats

Under general anaesthesia, a small burr hole is made in the animal's skull and a tiny micro-dialysis probe is implanted in the limbic region of the brain. The animals recover overnight and then are challenged with novel, naturalistic environmental stimuli over a period of 2 to 5 hours – e.g. being placed in a new cage, being exposed to a bright light, or another (new) animal of the same species and strain. The chemical effects of these stimuli on the limbic region of the brain are monitored by sampling via the micro-dialysis probe.

By these means, the changes in brain chemistry of animals that receive appropriate doses of drugs used to treat anxiety and depression can be compared with control animals that do not receive the drugs. In some studies, the responses of genetically altered (knock-out) mice are also compared. These knock-out mice show no overt phenotypic changes, except when challenged with novel environmental stimuli, when they behave like mice given anti-depressant drugs.

The pharmaceuticals employed are usually well established for clinical use in humans, and have the potential to alter the emotional state of the animals in a 'positive' way – e.g. reducing anxiety, enhancing feelings of well-being.

This technique is sensitive enough to detect small chemical changes in the brain of rodents following experimental challenges that cause no, or minimal, discomfort. Following the successful development of this technique in animals, it is being used increasingly in humans in a range of medical contexts (e.g. during surgery and for slow, localised infusion of medicines).

Severity scores:

Grid 1 – maximum severity – D – because of the surgery involved

Grid 2 – remaining severity – A – because the stimuli involved are mild (naturalistic) and brief, and the drugs administered do not induce adverse effects.

5: Protocol to study gene function in limb development (involving breeding, but not otherwise using, genetically altered mice)

This protocol covers the breeding of 15 lines of genetically altered mice used to study the genetic mechanisms involved in limb development. The mice so bred are killed by a Schedule 1 method at different stages (e.g. embryo, fetus, neonate etc), so that the effects of the genetic alterations on the developing limbs can be assessed *post mortem*.

Severity classification

In most of the lines that are maintained as breeding colonies under this protocol, the genetic alterations cause no discernible phenotypic effect (but the animals require DNA analysis from tissue taken from ear biopsy). Even if biopsy is not performed, such breeding is itself a regulated procedure under the Act and the number of animals involved must be reported to the Home Office as part of the annual Returns. The severity of the procedure would be classified as 'mild'.

In one of the 15 lines, the genetic change results in a limb defect, and, in another line, the development of smaller than normal ear pinnae. In these lines, ear biopsy is not required because the obvious phenotypes indicate the animals' genotypes. In both cases, the severity of the impact on the animals, in terms of pain, suffering, distress or lasting harm, is uncertain – but, giving the animals the benefit of the doubt, would be classified as 'moderate'.

On this basis, the severity categorisation using the two-grid scheme would be as follows for animals killed at 4 weeks (or later):

<u>For the phenotypically 'normal' lines killed at 4 weeks:</u>	CC
<u>For the 'limb defect' line killed at 4 weeks:</u>	FF
<u>For the 'small ear pinna' line killed at 4 weeks:</u>	FF

However, depending on the stage of development at which the animals are killed, other combinations would be required, to reflect the differences in 'duration' of the 'breeding procedures' involved. (Note that it is unclear what duration to allocate when an animal is killed at the embryo stage of development, nor how this would compare with duration allocated to an animal killed at the one-day neonate stage, for example).

Comment

It can be argued that the two-grid intensity-duration classification system brings unnecessary complications when used to report the severity of procedures involving breeding, but not otherwise using, genetically altered animals. When animals are killed at different times, a large number of code combinations are possible, because of the different 'durations' involved. But does it make sense to distinguish 'duration' of these procedures in this way, particularly for lines in which no overt adverse effect is caused by the genetic alteration?

In this light, it is suggested that the severity of adverse effects experienced by genetically altered animals bred but not otherwise used should be recorded in a different way, using a single code to reflect *only* the intensity of the effects of the genetic change, and not the duration. The numbers of such animals Returned could then be reported by severity category in a special table in the Home Office *Statistics*.

The protocol described above is relatively simple, in that the genetically altered animals are 'merely' bred and not used in other regulated procedures. It is noted that the number of possible severity code combinations is further increased for animals that are bred and then subsequently used in other regulated procedures.

6: Mouse mutagenesis study of a degenerative disease

This study aims to dissect the genetic influences on time of onset of a specific degenerative disease.

1. Male mice are injected with a chemical mutagen, which produces random mutations in their sperm. Approximately 50 different genes carry changes in each sperm.
2. The mutagenized males are mated with female mice that have been genetically altered to carry a gene for a late onset degenerative disease.
3. The resulting offspring are tested for presence of the transgene, by ear biopsy at 3 weeks of age. Because the mothers carry the transgene in heterozygous form, around 50% of the offspring will not bear the transgene and are therefore humanely killed.
4. The remaining offspring, which carry both the transgene and random mutations caused by the chemical treatment, are then screened for signs of onset of the degenerative disease, using a scoring scheme specific to the disease. The aim is to ascertain how the gene changes caused by the chemical treatment influence the time of onset of the degeneration (i.e. to dissect the role of gene modifiers in the onset of the disease).
5. Mice showing signs of degeneration are assessed using a specifically designed welfare assessment scoring scheme. Mice approaching humane end points (under HO guidance) are humanely culled.

Severity categorisation

(i) Experience of progeny that do not bear the transgene and are humanely killed

These animals are produced under a 'moderate' severity limit protocol, but do not bear the gene for the degenerative disease. It can therefore be argued that the maximum suffering they experience is that caused by the ear biopsy, which, under current HO guidance, would be classified as 'mild'. On this view, use of the two-grid scheme would lead to the severity classification AC for these animals.

However, it might also be argued that, although the effects are unknown, the mutation load borne by these animals by virtue of the ENU treatment of their fathers, should push the severity classification into the 'moderate' category, i.e. DF.

(ii) Experience of mice screened for neurodegenerative disease

The Working Group was presented with adverse effect data for a sample of ten mice screened for signs of degeneration. Of these, 8 out of 10 experienced mild effects (classified as described above) during the week after onset of the condition, and two experienced more moderate effects – one of which was humanely killed prior to exceeding its severity limit due to considerable weight loss over the week.

The 8 mice experiencing mild effects were classified as CC and the mice experiencing more moderate effects as FE.

7: Scoring studies to assess treatments of arthritic symptoms in mice

Study design

Collagen together with adjuvant is injected subcutaneously on a single occasion. This results in a model of arthritis 3 to 4 weeks later. Animals are dosed daily with compound or vehicle and the effect on inhibiting joint swelling and limping monitored for a further 2 to 3 weeks.

Protocol severity limit

This is “moderate” because of development of swollen joints and limping and possible side effect of small skin ulcers at the site of initial injection. Severity is limited by treatment of ulcers and killing the animals if they lose mobility or their general health begins to decline irreversibly.

Monitoring and assessment of severity of adverse effects

Animals are already examined daily and effects scored and recorded locally.

A 5 point scoring system is used for both arthritis and ulcers.

For joint swelling, the first signs of swelling in digits or ankle are “mild”; any greater swelling “moderate”. For ulcers anything up to 10mm² scab at the base of the tail is scored as “mild”; a larger scab or open wound, which occurs in < 7% animals, is scored as “moderate”.

Severity classification

Because of the lack of correlation between “arthritis” and development of ulcers each animal has to be evaluated individually for maximum severity (moderate arthritis or ulceration), and then assessed for remaining severity, looking at the overlap of effects and the time for which each lasted at each severity level.

If the proposed two-grid system was used the possible combinations and resulting severity scores would be: AA; CB; CC; EB; EC; FC; FE; FF.

In this example it is difficult to predict the severity category in advance for each grid and use of the scheme is a fairly lengthy process due to fact that wanted and unwanted adverse effects are recorded separately for each individual (as part of the protocol).

8: Three regulatory toxicity studies

In these examples, reporting was 'by group', because within each group individual animals experienced similar adverse effects (or lack of them). Had there been significant differences in effects on individuals within a given group, reporting would have had to be 'by individual'.

Study 1: Long-term study involving dogs

Dogs were dosed intravenously with a pharmaceutical compound at daily intervals, for 39 weeks. Controls were dosed with sterile saline. It was anticipated that the compound would produce only mild effects in the animals. The severity of any clinical signs observed was categorised using Company-wide guidelines (evolved over a period of time through discussion and practice).

The study design and resulting severity classifications using the two-grid system were as follows:

Group	Dose	Number of Animals	Severity classification using the two-grid scheme, taking 'mild' to refer to 'up to mild'
1	Control	8	All CC
2	Low	8	All CC
3	Intermediate	8	All CC
4	High	8	All CC

The severities of the adverse effects arising in each group were all classified as CC, taking the view outlined in section 7.1.2 of this report, that these categories encompass effects 'up to mild'. However, such a classification does not distinguish the different effects experienced by control *cf.* dosed animals, nor by low *cf.* higher dose groups. This raises the question of whether there should be a 'no effect' category to record animals showing no overt clinical signs (controls and low dose groups in this example).

Study 2: Short-term cyclical dosing study involving rats

This study had a more complicated design, involving dosing rats intravenously for five days with a pharmaceutical compound. In groups 1a and 4a, this was followed by a 6-week observation period, and in groups 1, 2, 3 and 4 by a 5-day observation period. Controls were dosed with sterile saline. It was anticipated that the compound would produce effects *up to moderate severity*. The severity of any clinical signs observed was classified using the guidance referred to above.

The study design and resulting severity classification using the two-grid system, again taking mild to refer to 'up to' mild, were as follows:

Group	Dose	Regime	Number of animals	Severity classification
1	Control	5 day IV + 5 days observation	30	CB
1a	Control	5 day IV + 6 weeks observation	20	CC
2	Low	5 day IV + 5 days observation	50	CB
3	Intermediate	5 day IV + 5 days observation	50	CB
4	High	5 day IV + 5 days observation	64	EA
4a	High	5 day IV + 6 weeks observation	20	FF

Study 3: Acute toxic class study involving rats (OECD Test Guideline 423)

This study involved giving groups of 3 rats a single oral dose of a pharmaceutical intermediate, followed by observation for up to two weeks. Doses were administered in a step-wise fashion, so as to enable the acute toxicity of the compound to be classified according to the regulatory guideline whilst using the minimum number of animals possible and causing the least possible suffering. Animals in Group 1 were humanely killed when substantial adverse effects became apparent. The severity of clinical signs observed was again classified using the guidance referred to above.

The study design and resulting severity classification using the two-grid system were as follows:

Group	Dose	Number of animals	Severity classification
1	High	3	GG
2	Intermediate	3	DB
3	Intermediate	3	DB

APPENDIX 3:

'Degrees of severity' categories, used for retrospective reporting of actual adverse effects in Switzerland²⁶

In Switzerland, at the end of each experiment, researchers are required to report the number of animals which suffered (or not) each of four different 'degrees of severity' of adverse effect, so that annual statistics can be compiled. The definitions of the severity categories (numbered 0 to 3) encompass intensity as well as duration of adverse effect. The basic definitions of the different 'degrees of severity' are listed below, and are elaborated in a detailed guidance document containing numerous examples¹¹.

- Degree of Severity 0:** For animals **not** exposed to **any** pain, suffering injury or severe fear through interventions and procedures for experimental purposes, and whose general well-being **was not significantly impaired**.
e.g. animals used for experiments which do not require authorisation, or control animals on which no interventions with adverse effects were conducted
- Degree of Severity 1:** Animals suffering a **mild (minor) temporary or brief adverse effect**.
e.g. animals killed in pre-terminal narcosis or rabbits immunised without use of Freund's adjuvant
- Degree of Severity 2:** Animals suffering *either*: a **moderate (medium-severe) short-term adverse effect** *or*: a **mild (slight) medium to long-term adverse effect**.
e.g. animals in which electrodes were implanted in the brain or those which underwent adrenalectomy
- Degree of Severity 3:** Animals suffering *either*: a **severe or very severe adverse effect** *or*: a **moderate (medium-severe) medium to long-term adverse effect**.
e.g. animals which underwent thoracotomy or ischaemia of the brain with subsequent substantial functional disorders. In long-term studies (e.g. lifetime toxicity study) spontaneously deceased animals must be allocated to severity degree 3, unless such animals are humanely killed at an earlier stage in clinical symptoms, when a correspondingly lower degree of severity will be allocated.

²⁶ The Box amalgamates descriptions from two sources: Swiss Federal Veterinary Office (1994). *Explanatory notes on interim and final reports of animal experiments*. <http://www.bvet.admin.ch/tierschutz/00237/00832/index.html?lang=en>; and Swiss Federal Veterinary Office (undated). *Classification of animal experiments according to grades of severity prior to the experiment (stress categories)*.

APPENDIX 4:

Suggested opportunities to reduce and/or simplify processes associated with A(SP)A if retrospective severity reporting is introduced

1 Processes and procedures directly impacted

- 1.1 Re-use – if prior authority to re-use animals has been granted the reporting process could be simplified. The two-grid system would report an animal's entire experience during its regulated use (taking into account cumulative effects).
- 1.2 Infringements – unintentional breaches of severity, such as an animal found dead on first daily inspection, could be reported in a consistent way using a retrospective system. This 'open' style of reporting fosters a responsible attitude, ensures feedback and learning within the establishment and works successfully in the Swiss system.
- 1.3 Genetically altered animals – the proposed system for retrospective severity recording presents difficulties for those animals that are phenotypically 'normal'. The UK requirements for reporting GM animals is more onerous than that required under European Legislation (Directive 86/609) and this is an opportunity to amend this requirement and simplify the process.

2 Processes and procedures benefiting from review

The following processes and procedures have been identified during the course of discussion in the working group or as a result of feedback from Project Licence holders during the consultation exercise.

- 2.1 Certificate of Designation – The Certificate Holder, not the Home Office, should be responsible for maintaining the inventory of animal rooms/procedure areas for specific purposes/species, which is then available for audit by the Inspectorate.
- 2.2 Personal Licences – (i) the simple categorisation (A,B,C) procedure as previously used should be re-introduced. The current system is over-complicated (more than 70 common codes in an establishment) and does not allow easy transfer of staff between establishments within the same organisation; and (ii) the addition of regulated procedures to a licence should be by a process of notification rather than authorisation using the establishment's evidence of training and competence.

3 Approval of Project Licences containing only mild procedures

Project Licences that contain only mild procedures should be approved within the establishment using the ERP and notified to the Home Office. No response within a given time (1 month) would trigger commencement of work (Tell and Wait). This notification procedure is current practice in the majority of EU member states.

4 Amendments to Project Licences

Amendments to protocols of a mild or moderate severity limit that do not alter the severity limit should be via a process of notification rather than authorisation (Tell and Do).