© 2015 Universities Federation for Animal Welfare
The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2015, 24: 139-149
ISSN 0962-7286
doi: 10.7120/09627286.24.2.139

Refinement of welfare through development of a quantitative system for assessment of lifetime experience

S Wolfensohn*†, S Sharpe‡, I Hall‡, S Lawrence‡, S Kitchen‡ and M Dennis‡

† University of Surrey, Faculty of Health and Medical Sciences, School of Veterinary Medicine, Duke of Kent Building, Guildford, Surrey GU2 7XH, UK
‡ Public Health England, Porton Down, Salisbury SP4 0JG, UK
* Contact for correspondence and requests for reprints: s.wolfensohn@surrey.ac.uk

Abstract

*This paper proposes a system that uses intrinsic study data to provide a clear visualisation of the stresses involved during the animal's life history that can be applied to all types of studies, even those not requiring invasive techniques. Thus, it provides an opportunity for researchers to identify and refine key events which impact on the welfare of an animal, and to explain clearly the totality of any necessary harms when justifying the research. Assessment of animal welfare depends on measurement of a number of parameters which will vary according to species, the animal's environment and the scientific procedure, all of which are inter-related. Currently, there are few tools to assess the effects of lifetime events on welfare or even, in some cases, to recognise that they have an impact on the level of suffering. A matrix to assess the combined effects of environment, experimental and contingent events on welfare has been applied, retrospectively, to programmes of work involving macaques (*Macaca mulatta *and* M. fascicularis*). Lifetime records, available for animals from their birth in the breeding colony through to experimental use in vaccine efficacy evaluation studies, were analysed as a robust validation test for the assessment matrix and refinement of the way in which information on these events is captured. A meaningful assessment method is required prospectively for project licence applications and retrospectively for licence review or decisions on re-use. The analysis will provide information that would support the application of refinements that would optimally enhance the lives of experimental animals.*

Keywords: *animal welfare, lifetime experience, macaque models of infectious disease, refinement of procedures, retrospective assessment, TB vaccine assessment*

Introduction

The EU Directive 2010/63 on The Protection of Animals used for Scientific Procedures encompasses the concept of cumulative severity, in which the whole experience of each animal is taken into account when assessing the severity classification of a programme of work. This is a variation from previous classification systems where there was a requirement only to consider the direct suffering caused by a particular technique rather than taking into account any contingent events.

Specifically, the Directive requires "...taking into account the lifetime experience of individual animals..." (paragraph 25), "...to enhance the lifetime experience of the animals..." (paragraph 31), "...to reduce the duration and intensity of suffering to the minimum possible..." (Article 13.3b). It requires that the severity category shall take into account the nature of pain, suffering, distress and lasting harm, and its intensity, the duration, frequency and multiplicity of techniques, and the cumulative suffering within a procedure (Annex VIII). Annex VIII contains guidance on assignment criteria when considering cumulative suffering within a procedure but the examples are a mix of 'techniques' and 'protocols'. They are not based on the use of specific refinement measures which can have a significant impact on the actual severity experienced by the animal.

The level of suffering experienced by an animal is the combination of direct effects on welfare (the procedural protocol on a licence in terms of the actual procedure and the combination of techniques and their subsequent outcomes) and any clinical condition from which the animal suffers which may not be due to the procedure (a bite wound or bullying by conspecifics, for example), plus any contingent effects related to housing, husbandry or transportation. The duration of each of these and the intervals between events must also be taken into account and the extent to which an animal is deviating from normality (Morton & Griffiths 1985).

There is an increasing interest in developing methods of assessing the lifetime experience of experimental animals by addressing issues such as the cumulative effect of a

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

series of procedures and the retrospective assessment, on completion of a study, of *actual* severity experienced by individual animals. The EU Directive 2010/63 requires that retrospective assessment shall be carried out:

> to evaluate whether the objectives of the project were achieved, the harm inflicted on animals, including the numbers and species used and the severity of the procedures used and any elements that may contribute to the further implementation of the requirement of replacement, reduction and refinement (article 39).

The Working Document on a Severity Assessment Framework under the EU Directive (2012) requires recording of the effects of procedural events for retrospective reporting of actual severity at the end of procedures, but it has been recommended that retrospective assessment should be based on the continuous collection of data as experiments progress. The UK Home Office now requires that data are collected on all animals involved in regulated procedures completed, including actual severity, and these data will be required to be submitted on an annual basis from January 2015. The Animal Procedures Committee (2013) recommended collecting data to assess an overview of the animal's lifetime experience with key events and quality of the environment, including benefits of any refinements that have been developed and a log of adverse events (non-procedural, generic and intended effects of the procedure and complications) including their impact on welfare.

With this in mind, the system detailed here has been developed to allow recording and assessment of lifetime events for individual animals (Honess & Wolfensohn 2010). In order to test the suitability of this system, in terms of ease of use and discrimination of differences in lifetime experience, it has been tested using data from a series of experiments involving a macaque model of tuberculosis.

The macaque model of tuberculosis

Tuberculosis (TB) is a major health problem, especially in low income countries. There are 1.4 million deaths per year from TB (WHO 2014) and it is estimated that a third of the world's population is latently infected. The devastating effects of TB infection have been exacerbated by the emergence of multi-drug resistant strains and HIV co-infection. There is an urgent need for improved interventions; including a new vaccine as the only currently available vaccine, BCG, offers limited protection. Macaque studies are being performed in parallel with human clinical trials to assess the utility of a challenge model as macaques offer the opportunity to compare the BCG and M.tb challenge models and validate a BCG challenge model. The work presented here has used data from completed studies on macaques to refine and demonstrate utility of a lifetime assessment system previously described by Honess and Wolfensohn (2010).

Since the establishment of the macaque aerosol challenge model of TB, the facility at Public Health England (PHE) has worked to introduce new practices to improve the welfare of animals under study that have included changes in housing. The refined assessment system was used to evaluate the lifetime experience of animals in studies to evaluate new TB vaccines, and the lifetime experience of animals in studies benefiting from these new approaches were compared to determine the ability of the system to provide a measure capable of quantifying the impact of key changes in housing or experimental practices. The acquisition of two adjacent macaque breeding colonies during the lifetime of the project has provided a unique opportunity to test potential refinement strategies afforded by having both breeding and experimental facilities on the same site.

Materials and methods

All data used in this report were obtained retrospectively from completed studies where the findings have been previously reported (Sharpe *et al* 2009, 2010). Although lifetime data were available for most of the animals in these studies a specific time-period was selected that was comparable for each study so that the ability of the welfare assessment system to differentiate housing, environmental and experimental experiences could be demonstrated. The data that were entered for each individual animal were based on information recorded on each occasion that a procedure or other intervention took place.

Experimental animals

With the exception of one study (study 2), all animals described in this paper were rhesus macaques (*Macaca mulatta*) obtained from a long-established UK breeding colony. For comparative purposes, historical data from study 2 involving cynomolgus macaques (*M. fascicularis*) imported from a Home Office approved non-UK source (the Guangxi Grand Forest Scientific Primate Company) are also reported. All studies were conducted under project licences approved by the Ethical Review Process of PHE, Porton, Salisbury, UK and the Home Office, UK. Animals were housed according to the Home Office (UK) Code of Practice for the Housing and Care of Animals Used in Scientific Procedures (1995) and following the National Committee for Refinement, Reduction and Replacement (NC3Rs) Guidelines on Primate Accommodation, Care and Use (NC3Rs 2007). Throughout the course of these studies two types of caging systems were used at biocontainment level 2 (CL2) and two versions at biocontainment level 3 (CL3). These cages differed in the amount of height, balcony access and opportunity for enrichment that was provided. In the case of CL2 caging, the newer style provided balcony access, more provision for 3D enrichment and the ability to forage in deep litter (manipulable enrichment). Similarly, the later version of CL3 accommodation provided greater height and better opportunities for 3D and manipulable enrichment. These differences are summarised in Tables 1 and 2. The programmes of work evaluated were all designed to examine the efficacy of vaccine regimes against TB, where the efficacy of new vaccine candidates is compared to the efficacy of BCG. These involved a period of vaccination followed by either a challenge with M.tb or with BCG as a surrogate challenge.

© 2015 Universities Federation for Animal Welfare

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

Table 1 Summary of the four studies assessed retrospectively.

Study number	Species and origin	Vaccination phase: Containment level and Housing	Challenge phase: organism and dose range	Group	Challenge phase: Containment level and Housing	Combined welfare assessment score for post challenge phase	Time period post challenge phase to efficacy
1	Rhesus, UK	Experimental facility: CL2v1, linked cages	*Mycobacterium tuberculosis*, 800–1,200 CFU	BCG vaccinated	CL3v2, until week 28 post challenge, then 13 weeks in CL3v1, then back to CL3v2	22.9	52 weeks
				unvaccinated	CL3v1, final 5 weeks one animal in CL3v2	24.3	52 weeks
2	Cynomolgus, China	Experimental facility: CL2v2, balcony cages	*Mycobacterium tuberculosis* 7,000–10,000 CFU	BCG vaccinated	CL3v2	18.5	26 weeks
				unvaccinated	CL3v1	19.2	26 weeks
3	Rhesus, UK	Breeding facility	*Mycobacterium tuberculosis* 500–1,000 CFU	BCG vaccinated	CL3v1	24.4	16 weeks
				unvaccinated	CL3v1	27.0	16 weeks
4	Rhesus, UK	Experimental facility: CL2v2, balcony cages	BCG	BCG vaccinated	CL2v2	10.3	2 weeks
				unvaccinated	CL2v2	10.5	2 weeks

Table 2 Scores for the different environments at PHE.

		Containment level 2		Containment level 3	
	Breeding colony	**CL2v1: Older style with upper extensions**	**CL2v2: New style with balcony**	**CL3v1: High containment**	**CL3v2: High containment**
Housing	2	5	3	6/7/8*	4
Group size	1	3	3/4	3/4/6	3
3D enrichment	1	4	3	6/7/8*	5
Manipulable enrichment	1	6	2	7	6
Average score	1.25	4.50	2.75/3.00	5.50/6.25/7.25	4.50

* Scores depend on the weight of the animal: Lowest score < 4 kg, intermediate score 4–6 kg, highest score > 6 kg.

Vaccination

Animals were immunised intradermally in the upper left arm with 100 µl BCG, Danish strain 1331 (SSI, Copenhagen, Denmark) prepared and administered according to manufacturer's instructions for preparation of vaccine for administration to human adults.

M. tuberculosis challenge procedures

Animals were challenged by aerosol inhalation of M.tb (Erdman strain) as previously described in Sharpe *et al* (2009). Animals received doses ranging between 800–1,200 CFU (study 1), 500–1,000 CFU (study 3) and 7,000–10,000 CFU (study 2).

M. tuberculosis challenge study design (Studies 1, 2, 3)

Six macaques in each study received an intradermal vaccination with BCG. Prior to challenge, animals in studies 1 and 2 were housed in the experimental facility under CL2 conditions, animals in study 3 were housed within the breeding facility in which they were born. Twenty-one weeks after immunisation, the BCG vaccinated group, together with four (study 1) or six (studies 2 and 3) naïve controls received an aerosol challenge with M.tb. Following challenge, animals were housed for a maximum of 52 weeks (study 1; Sharpe *et al* 2010), 26 weeks (study 2) or 16 weeks (study 3) in the biocontainment facility under CL3 conditions (Dennis 2010) and

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

Figure 1

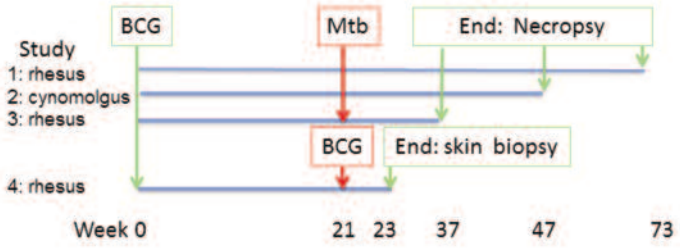


Diagram to show the timelines of studies 1-4.

monitored daily for behavioural changes and clinical signs of disease. Clinical parameters such as weight and body temperature were measured and blood samples collected for evaluation of immune responses, red blood cell haemoglobin levels and erythrocyte sedimentation rate, at two weekly intervals throughout the study.

BCG 'challenge' study design (Study 4)

Twenty-one weeks after BCG immunisation, six BCG-vaccinated animals together with six non-vaccinated animals were 'challenged' with BCG via the intradermal route. Following challenge (C0), blood samples were collected on days C+2, C+7, and C+14 for immunological studies and measurement of BCG burden. A skin biopsy was collected from the BCG challenge site 14 days after challenge for estimation of BCG burden. Clinical parameters were measured in all animals each time animals were sedated during the study period. Animals were housed throughout the study in the experimental facility under CL2 conditions. Figure 1 and Table 1 show the timelines of studies 1–4.

Description of the Extended Welfare Assessment Grid (EWAG) system

From the individual records maintained at the time for each animal, a welfare assessment score was applied at each moment in the animal's life when there was a record of some intervention having taken place. This system scores four parameters (Physical, Psychological, Environmental, and Procedural [experimental and/or clinical]). These four parameters together contribute to the level of combined severity and are scored to obtain a quantitative measure. Within each parameter various factors are scored between 1 and 10 where a score of 1 indicates the best possible state (lowest possible impact on welfare) for the respective factor, whilst a score of 10 would be the worst possible state (highest possible impact on welfare). For each parameter the factors scores were averaged to obtain a figure for the parameter reflecting its contribution to severity and thus impact on quality of life at that time-point. The four parameters form a 2-dimensional grid. When the assessment of the animal is made at a given time-point and the relevant scores are marked on the arms of the x and y axes, these points are joined together to create a two-dimensional polygon the area of which can be calculated by using standard geometry and the value derived may be seen as a numeric representation of the welfare condition of an animal, or group of animals, at one moment in time — the combined welfare assessment score (CWAS).

Notice that the four parameters are not independent of each other. For example, an animal that undergoes a procedure may well have a change in routine that affects its use of enrichment, and its activity level may be affected by being put into restricted housing. This would be reflected in the scores for all four parameters. The lack of independence between the factors and their consequences is reviewed by Collins (2012) who concludes that this does not affect the scoring of individual-focused factors of intensity or duration of adverse experience. Indeed, it may be argued that if the actions of a hazard lead to multiple types of consequences, then standardising will mean that the risk outcomes for this hazard could be underestimated (Collins 2012). Intensity and duration of consequences are not independent since an individual's level of adverse experience will be a function not only of current intensity, but also how long it has been attempting to cope with the insult, and whether and to what extent this has impacted on immune, endocrine and nervous-sensory systems (Broom & Johnson 1993). When evaluating an animal's welfare it is necessary to consider the timing, duration and frequency of welfare assessment measures which will take into account the species/strain of animal, the experimental design, the housing environment, husbandry practices and the animals' normal circadian rhythm. Care must be taken when interpreting numerical scores of welfare assessment since, for example, a score of three for staring coat and three for isolation from conspecifics does not equate to a score of six for a distended bowel. A score of six is not suffering twice as much a score of three (Hawkins *et al* 2011).

Whilst there are certain fundamental factors that are relevant to all animals undergoing experimental use, there are obviously decisions to be made on which factors are to be scored and the relative weight that is given to them. The system described is highly flexible in this respect and should be tailored to each individual type of experiment. Thus, experiments on a pulmonary infectious disease model, such as tuberculosis, will assess relevant factors that reflect the animals' condition (such as respiration rate, level of dyspnoea). When used to assess other types of experiments, for example in neuroscience, the factors should reflect conditions that may occur in such procedures (such as withholding

© 2015 Universities Federation for Animal Welfare

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

access to fluids, surgical intervention, incidence of seizures etc). Thus, careful consideration must be made on which criteria are to be scored.

For the specific scores for each factor, some examples are given below. However, all these elements will need to be scored with advice and training from the attending veterinary surgeon and experienced animal care staff, since they require a level of experience in clinical assessment. A baseline of good welfare should first be defined for the particular species and has been defined as:

> the state of being in (ie impact on) animals when their nutritional, environmental, health, behavioural and mental needs are met (Hawkins *et al* 2011).

The potential causes of deviation from the ideal welfare state may include inappropriate housing or healthcare and the effect of procedures; but the key is to select appropriate welfare indicators that are relevant for the species and give a general indication of welfare and those specific to the project (for more detail, see Hawkins *et al* 2011). Once they have been agreed and ascribed relevant to the effects of a specific study and husbandry system, they can be applied by research and care staff. For more detail on welfare assessment scoring see Wolfensohn and Honess (2005). It is not the purpose of this paper to give a set of values for all of the criteria, rather it is up to research teams to agree this based on experience and discussion.

For this report the parameters were scored as follows.

Physical

Factors

General condition

Weight-loss, condition score etc (see Wolfensohn & Honess 2005; chapter 5). Weight loss of zero up to 1.9% score 2, from 2–3.9% score 3, from 4–5.9% score 4, from 6–9.9% score 5, from 10–14.9% score 6, from 15–19.9% score 7, from 20–24.9% score 8, over 25% score 10. In growing animals it is important to score against expected weight gain according to the normal growth curve.

Clinical assessment

For example, cough, diarrhoea, ascites, tachypnoea (rapid breathing), dyspnoea (struggling or irregular breathing), vomiting, temperature changes). It is necessary to consider how much the clinical condition is affecting the animal (ie how ill is it) not just the extent of the clinical sign. For example, the animal may have an increased respiration rate after exercise but not be stressed by it or it may have an increased respiration rate due to infection which is causing it distress. These would be scored at different levels. Also, it is possible to give the same score for different clinical conditions (eg diarrhoea and dyspnoea) if they are affecting the animal's welfare to the same extent. If there are a combination of conditions that are occurring at the same time it is the net effect on the animals that needs to be reflected in the score.

Activity level, mobility

In addition to reduced activity, note that hyperactivity may be abnormal and result in poor welfare, incurring a score. The absolute activity level will also need to take into account the age of the animal since juveniles will be more active.

Presence of injury

If there is an injury then a score will depend on the level to which it affects the animal's welfare. Absence of any injury scores 1, a serious debilitating injury would score 9 or 10.

Not eating/drinking

Normal intake score 1, food/water intake less than 40% for 3 days score 6, food/water intake below 40% for 7 days or anorexic for 3 days score 9. This should also include withholding of food/water so that intake is below normal levels.

Psychological/Behavioural

Factors

Stereotypy, self-harming, unusual grooming

The frequency and time spent in the abnormal behaviour, not just the extent of physical damage to the animal should be taken into account in the scoring. A behavioural ethogram can be used, such as in Wolfensohn and Lloyd (2013).

Response to catching event

Recognise if the animal is well trained and habituated, or stressed and frightened, or aggressive. The method used, for example, catching net, pole and collar, crush back cage or use of positive reinforcement training and how long it takes to catch the animal should be reflected in the score.

Hierarchy upset/dispute, aggression/bullying

Take into account the extent, duration and outcome.

Alopecia score (see Honess et al 2005)

If not recorded, do not score but mark as not applicable.

Use of enrichment

There may be lots of enrichment provided, but the score should reflect if the animal is using it.

Aversion to 'normal' events

This includes events such as staff interaction, cage cleaning etc. The score should reflect if the animal is well trained and habituated, or stressed and frightened, or aggressive.

Environmental

Factors

Housing

Free-range, zoo-type environment score 1. Consider the space provision, lighting, ventilation, weather exposure, animal-friendly materials, noise. Adjustment may be needed to take account of the species and size of the animals.

Group size

Singles score 10, pairs score 6, small groups (3 to 5) score 4, medium size groups (6 to 8) score 3, breeding groups score 1.

Provision of 3D enrichment

Consider the ability to climb, jump, hide, establish and maintain social hierarchy without aggression and fighting.

Provision of manipulable enrichment

Consider forage material, food provision and the variety of what is offered.

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

Table 3 Worked examples for routine licensed procedures on this study.

	Computed Tomography (CT) scan	Broncheo-alveolar Lavage (BAL)	Blood sample with sedation	Blood sample without sedation
Restraint	4	4	4	6
Sedation	4	4	4	1
Procedural/intervention effect	3	4	3	2
Daily routine changes	5	5	5	3
Average score	4.00	4.25	4.00	3.00

Contingent events

Consider if the animal has been moved, ongoing building works etc.

Scores for the different environments at PHE

See Table 2.

Procedural (experimental and/or clinical)

Factors

Restraint

The score should reflect the method used and the effect on the animal, such as use of positive reinforcement training, or if the animal is aggressive and frightened.

Sedation

Consider duration, quality of induction and recovery, effect on food intake and behaviour.

Effect of intervention

Irrespective of whether the intervention is a planned licensed procedure or a veterinary or husbandry procedure, assess to what extent this event has impacted on the animal's welfare at that time, even if done for the animal's benefit in the longer term.

Change in daily routine

Consider such things as withholding enrichment or food, restricting access to the usual living area as might be done before or after a planned surgical event, for example.

Worked examples for routine licensed procedures on this study

See Table 3 (as well as Figure 2 [see supplementary material to papers published in *Animal Welfare* on the UFAW website: www.ufaw.org.uk] for a scoring sheet as used in the facility).

As an example of how this is used for an individual animal (S36) see Figure 3 (see supplementary material to papers published in *Animal Welfare* on the UFAW website: www.ufaw.org.uk), using information taken from the Excel spreadsheet used to enter the scores. This animal was on study 3 in the unvaccinated group. The database was originally completed by an animal care technician and then re-assessed by a veterinary surgeon. As can be seen from the spreadsheet the key significant events in this animal's experimental life were moving into the containment facility and the development of clinical signs of tuberculosis. Using this animal as an example, a weight loss of up to 2% was given a score of 2 and from 2–4% a score of 3. The clinical score was increased to 5 (day 1,060) due to observation of respiratory signs that were clearly attributable to the progression of TB disease but were only seen under sedation; if they had been present in the animal when un-sedated they would have scored 7 or 8. The use of enrichment and aversion to normal events increased from 1 to 6 (day 1,018) when transferred from the colony to the experimental facility but went down from 6 to 4 or 5, respectively, (day 1,025) as the animal was judged to have become accustomed to its environment in the biocontainment facility. Similarly, the score for change in routine from 5 up to 7 and then down to 5 again is explained as the animal was moved with its conspecifics from the breeding colony into an experimental environment (days 967 to 1,025).

Results

Calculation of lifetime experience

The scores for each factor were entered into an Excel spreadsheet that automatically calculated the average score for the four parameters and quantified the welfare experience of the animal by calculating the area of the welfare assessment grid (WAG). A graph was then plotted to show the total lifetime experience welfare score (combined welfare assessment score CWAS) throughout the lifetime of each animal — see Figure 4 for an example for animal S36 from study 3 unvaccinated group. At any time-point on this graph a welfare assessment grid can be plotted to show the effects of each of the four parameters at that moment in time, the WAGs at the time-points day 801 (pre-challenge) and day 1,084 (7 weeks post-challenge). For animal S36 it can be seen in Figure 4 that the effect of the environment was significant in terms of its negative welfare state, in addition to the expected physical effects of the procedure of infectious challenge, and this allows for refinement to be tailored in order to have a maximal effect on improving the animals' welfare.

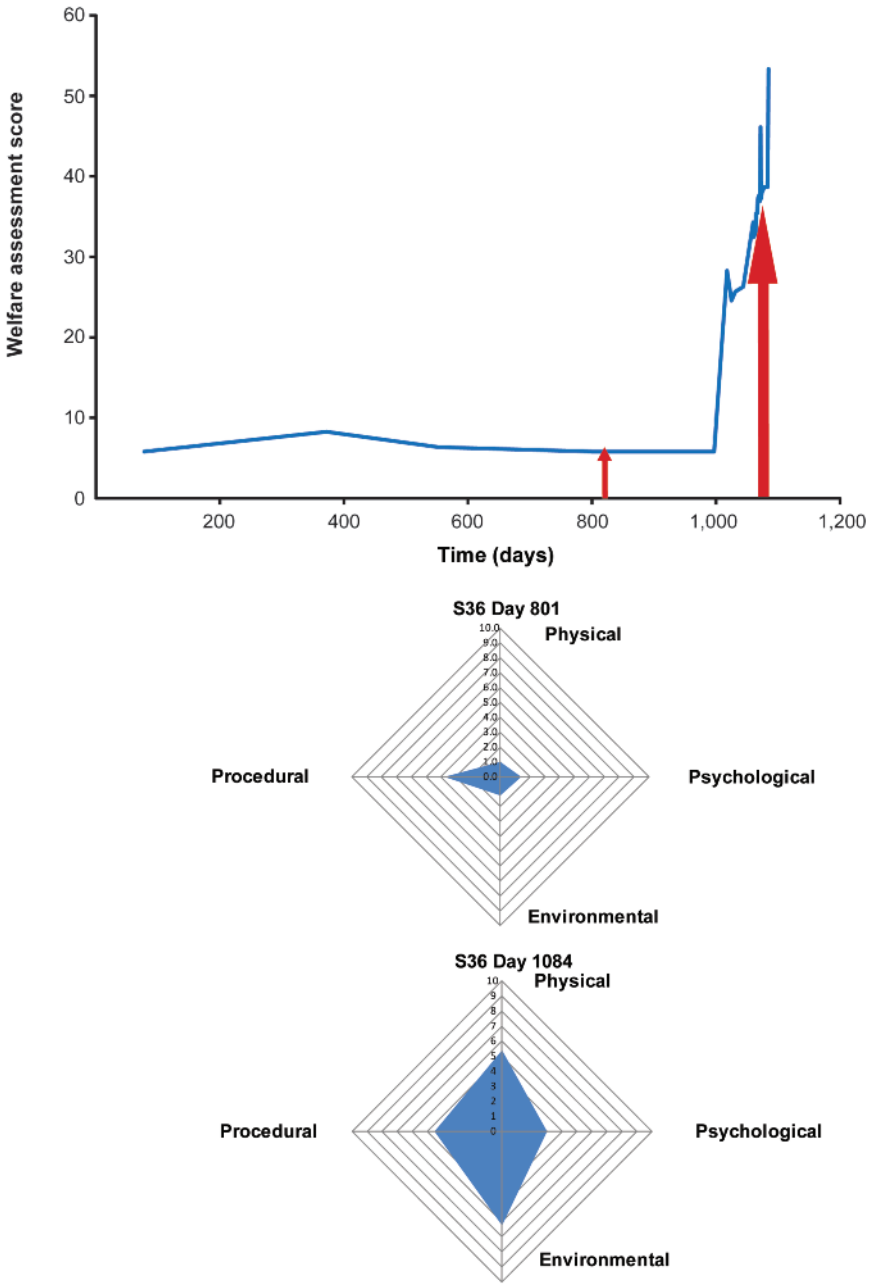
Use of the WAG to compare studies and demonstrate the effect of refinement of study design, environment and use of end-points

The WAG can be utilised to examine and quantify the effect on welfare following any changes in experimental design and environments, demonstrating the overall improvement these make to total lifetime experience. This will assist with the quantification of the 'harms' to the animals of studies to

© 2015 Universities Federation for Animal Welfare

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

Figure 4



Combined welfare assessment score throughout animal S36's life (Study 3 unvaccinated).

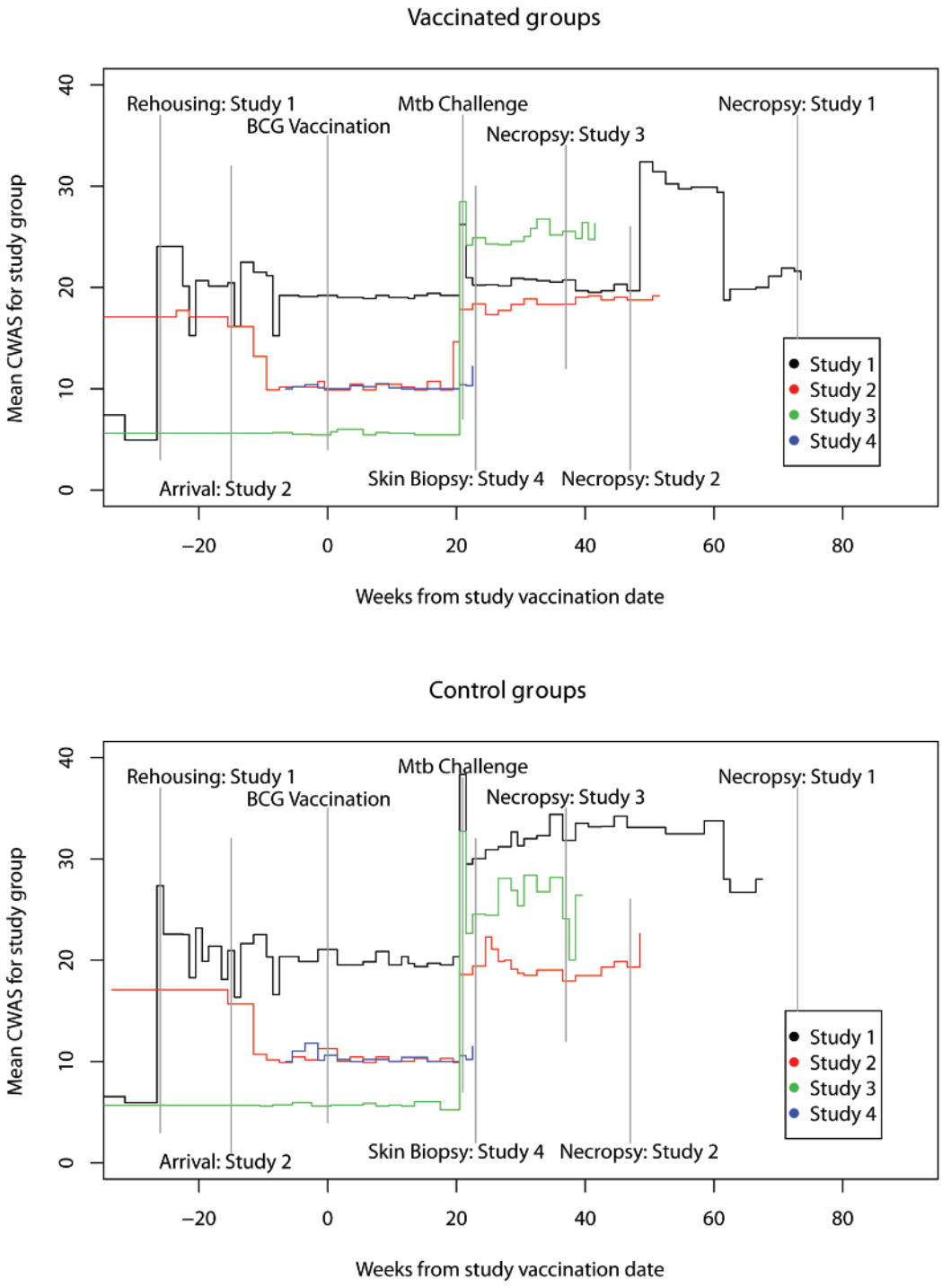
evaluate the vaccine efficacy. The WAG was applied retrospectively to quantify the lifetime experience of animals in four studies designed to evaluate the efficacy of new TB vaccines performed over a ten-year period at PHE (Table 1).

Although the intention is for this assessment system to be used in real-time from birth to death, we appreciate that this is most practicable when an establishment has both breeding and experimental facilities and therefore recognise that if animals are purchased from elsewhere the assessment can probably only start at the time of arrival. With this in mind, and in order to optimise the comparative power of this retrospective application of the system, when comparing groups of animals from these studies the combined welfare assessment score (CWAS) was averaged over a week's activity as well as across the group. Rather than use date of birth as a common reference point for each animal in a group, the week of common procedure (say, aerosol challenge or vaccination) was used as an anchor-point to enable simple visual comparison. Figure 5 shows the combined welfare assessment plots indicating the lifetime experience for all four studies for the vaccinated groups and the unvaccinated groups.

These results were used to determine the impact of study design. Specifically as follows.

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

Figure 5



Comparison of the lifetime experiences (CWAS) of BCG vaccinated and unvaccinated animals in studies 1 to 4.

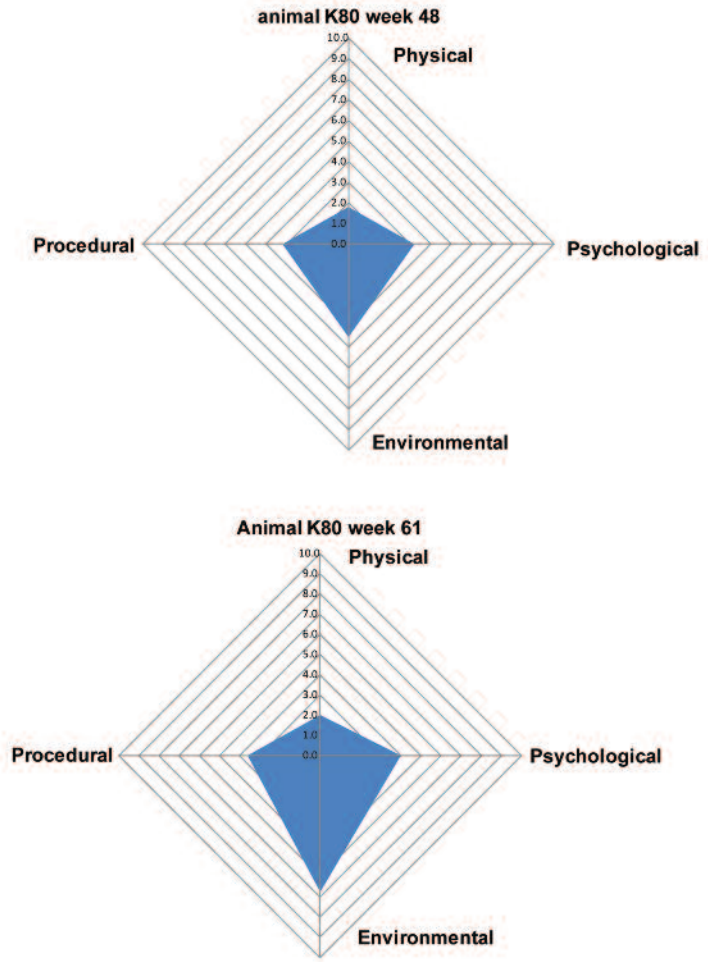
Design of the housing environment

By comparing the CWAS for the period from study enrolment before BCG vaccination through to challenge for all studies, the welfare effect of the different housing environments was demonstrated. In study 1 animals in the pre-challenge phase were housed in the first the version of group CL2 housing in the experimental facility, in studies 2 and 4 the animals were housed in refined CL2 housing in the experimental facility, and in study 3 the pre-challenge phase was conducted within the home breeding facility implementing a significant refinement. The analysis of the CWAS allows the quantification of the impact of housing facility on life experience. In Figure 5 it can be seen that the CWAS for vaccinated animals on study 1 shows a marked

© *2015 Universities Federation for Animal Welfare*

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

Figure 6



Animal K80 WAGs at weeks 48 and 61 to show effect of environment when transferred for 13 weeks to different type of BSL3 cage with poorer environment due to maintenance requirements.

increase between weeks 48 and 61. Examination of the individual WAGs for this period (see Figure 6) shows that the increase is due to a change in the environmental score. This was due to the animals having to be moved from more-refined CL3 housing to less-refined CL3 housing for 13 weeks while maintenance work was carried out.

The efficacy measure employed necessitating the use of either the survival or a fixed end-point, and the effect of duration of fixed end-point by comparing studies 1, 2 and 3

Comparison of the scores attributed to the animals following challenge with M.tb provided the opportunity to quantify differences in the lifetime experience when a survival end-point (study 1) was used, as opposed to a fixed end-point (studies 2 and 3) for the evaluation of vaccine efficacy. Study 1 had a 52-week post challenge follow-up period which was the amount of time required to obtain sufficient data to evaluate vaccine efficacy in this study design, whereas study 2 had a fixed end-point design of 26 weeks and study 3 had a fixed end-point design of 16 weeks as the post-challenge period. The combined effect on welfare of evaluation of vaccine efficacy can be calculated looking at the period after challenge to study end-point (52, 26, or 16 weeks) in all studies as shown in Table 1 and in Figure 5 and demonstrates the reduction in welfare 'cost' with decreasing the time post challenge.

The effect of challenge dose of M.tb used

Rhesus macaques on study 1 received a higher aerosol dose of M.tb than those on study 3. Comparison of the CWAS for the unvaccinated groups during the first 16 weeks after challenge in these studies when both were housed in the same environment shows the adverse effect of the increased dose on the animals in study 1.

The effect of vaccination to quantify the additional 'cost' of being in an unvaccinated control group animal in studies requiring aerosol challenge with M.tb

This can be seen by comparing the CWAS for BCG vaccinated groups of animals against the unvaccinated control groups within studies 1, 2 and 3 (Figure 5) where there is a trend for the scores to be higher in the unvaccinated groups.

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

The difference between vaccinated and unvaccinated groups in study 1, between weeks 21 and 48 after challenge, was particularly striking, however further interrogation of the individual WAGs for this period revealed that a portion of the reduction in score of the vaccinated group was due to the environment and the benefits afforded by the more refined CL3 housing in which they were maintained.

The effect of importation and use of species with different susceptibility and ability to control TB disease progression

By comparing the CWAS of animals on study 2 (cynomolgus macaques imported from China) with those on studies 1, 3 and 4 (home-bred rhesus macaques), the adverse effect of importation on the pre-challenge period can be seen. However, there is also a species difference in the susceptibility to M.tb which can be seen by comparing study 2 with studies 1 and 3 in the first 16 weeks post-challenge phase. Although the rhesus macaques were exposed to lower doses of M.tb than the cynomolgus macaques, they were more susceptible to the adverse effects associated with onset of disease.

The effect of changing from use of a pathogenic challenge organism to a non-pathogenic methodology and the duration of follow up

The impact on lifetime experience that could be achieved through development of a challenge system that did not require infection with M.tb and the consequent development of TB-induced disease, was evaluated through comparison of the CWAS during the post-challenge phase needed to evaluate vaccine efficacy. In study 1 (52 weeks after M.tb challenge) and study 2 (26 weeks after M.tb challenge fixed end-point) and study 3 (16 weeks after M.tb challenge fixed end-point) the CWAS was much greater compared to study 4 (2 weeks after BCG challenge).

Animal welfare implications

This work has provided a unique opportunity to refine and test the Extended Welfare Assessment Grid (Honess & Wolfensohn 2010) for its ability to assess the lifetime experience of non-human primates used in infectious disease research projects. Numerous methods of assessment of animal welfare have been reported but the majority only look at the animal at one moment in time. The Extended Welfare Assessment Grid represents a valuable tool to reflect the temporal component of suffering that is often overlooked and allows assessment of any suffering imposed by a combination of events that occur during the lifetime of an animal. The Grid examines the welfare of the animal at key points throughout its life taking into account the duration as well as the intensity of suffering producing numeric, as well as visual, representation of the animal's overall quality of life (FAWC 2009). It has been developed to monitor the lifetime experience and evaluates the physical condition, the psychological condition, the quality of the environment and the impact of procedural intervention and clinical conditions on the animal. The Extended Welfare Assessment Grid has particular value in producing an objective visual illustration of lifetime welfare status and is used in this paper to retrospectively review refinements in husbandry and study design to demonstrate objectively the effect on animal welfare. We have developed an assessment form (see Figure 2; www.ufaw.org.uk) that can be easily completed by animal carers or licence holders and will allow data entry to the welfare score system. It is important to recognise that the factors for each parameter can and should be modified to suit different types of experiments other than infectious studies. This can be done such that when data are entered in real time as an experiment progresses, its use enables those involved in the conduct of animal studies to plan and intervene with additional environmental enrichment, alterations to housing and husbandry practices, giving suitable treatments or carrying out euthanasia at an appropriate time, such as to limit the negative effect of interventions. Whilst this system may be seen to have particular value for long-term experiments, we have also demonstrated its utility for application to short- and mid-term programmes of work. Because of the design and nature of the programmes of work used here to test the system there were no signs of 'additive' effects as the experiment progressed. We recognise, however, that in some experiments the effects of one procedure or contingent event may compound those of a subsequent one.

This study has also been an opportunity to quantify the improvements that have been made to the lifetime experiences of animals in studies to evaluate new interventions against human diseases, through changes in study procedures and housing arrangements. Such a system will be invaluable in helping to make decisions at ethical review body discussion on harm:benefit balance and in considering the justification for carrying out the work. We recognise that as the scoring parameters are likely to be differently weighted according to the scientific field of the work, the system will not offer a direct comparison between one research programme and another. The strength of the system is to be able to evaluate where improvements can be made within a study or programme of work and we have demonstrated that the system allowed areas of environment, practices, contingent events and experimental design to be identified as the particular cause of any change in welfare, enabling refinements to be focused appropriately to maximise improvements in welfare wherever possible and this will assist study design in the future.

Conclusion

Ethical justification for carrying out a scientific procedure will be a balance between the harm to the animals and the benefit to society from the knowledge gained. The level of harm will be affected by the degree of implementation of the 3Rs (replacement, reduction and refinement). Thus, the overall level of severity will be directly affected by how the work is conducted and this will include elements of contingent suffering and direct suffering caused by procedures, such as sampling, administration of substances, surgery and the induction of disease. The use of this system allows the critical evaluation of the animals' quality of life and the recognition of signs of poor welfare, such that improvement strategies may be selected and implementation of the refinement loop (Lloyd *et al* 2008) will then assist in reducing the overall level of severity.

© *2015 Universities Federation for Animal Welfare*

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press

The general requirements for effective welfare assessment have been agreed as (Hawkins *et al* 2011):

- A team approach;
- Appropriate welfare indicators;
- A sound understanding of good welfare and the normal animal;
- Full recognition of all potential adverse effects;
- Consistency for all species;
- Consistency between observers;
- Appropriate recording systems.

While all of these requirements can be met for a particular moment in time, there is the need to assess the animal's whole quality of life (FAWC 2009) over the duration of the study in order to truly reflect its welfare. This project reviewed the welfare records to see how well the adverse effects had been predicted, recognised and alleviated; and to study the effect of refinements of housing and study design. The information gained was useful in liaising with the institutional animal welfare and ethical review body to seek guidance on the continued acceptability with respect to the local culture of care at the establishment. The Concordat on Openness on Animal Research (2012) highlights the importance of transparency about the actual level of suffering experienced by animals undergoing procedures.

Whilst we acknowledge that there is a considerable amount of work and discussion involved in setting up the parameters for assessment, in our experience, once set up, the system was user-friendly and easily applied by animal carers and technical staff. We are currently developing a database that will simplify data input and maximise the information that can be obtained by interrogation of the system.

References

Animal Procedures Committee 2013 Review of the assessment of cumulative severity and lifetime experience in non-human primates used in neuroscience research. *Report of the Primate Subcommittee of the Animal Procedures Committee*. https://www.gov.uk/government/publications/animal-procedures-committee-cumulative-severity-review

Broom DM and Johnson KG 1993 *Stress and Animal Welfare*. Kluwer Academic Publishers: Dordrecht, The Netherlands

Collins LM 2012 Benefits and pitfalls of welfare risk assessment. *Animal Welfare 21(S1)*: 73-79. http://dx.doi.org/10.7120/09627286.21.S1.73

Concordat on Openness on Animal Research 2012 www.understandinganimalresearch.org.uk

Dennis M 2010 Special housing arrangements. In: Kirkwood JKK and Hubrecht R (eds) *UFAW Handbook on the Care and Management of Laboratory and Other Research Animals, 8th Edition*. Wiley Blackwell: London, UK. http://dx.doi.org/10.1002 /9781444318777.ch11

European Directive 2010 *2010/63/EU*. http://ec.europa.eu/environment/chemicals/lab_animals/legislation_en.htm or http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:en:PDF

EU Directive 2012 *Working document on a severity assessment framework (2012)*. ec.europa.eu/environment/chemicals/lab_animals/pdf/Consensus%20doc%20on%20severity%20assessment.pdf

FAWC 2009 *Farm Animal Welfare in Great Britain: Past, Present and Future*. Farm Animal Welfare Council: London, UK. www.fawc.org.uk

Hawkins P, Morton DB, Burman O, Dennison N, Honess P, Jennings M, Lane S, Middleton V, Roughan JV, Wells S and Westwood K 2011 A guide to defining and implementing protocols for the welfare assessment of laboratory animals. Eleventh report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working group on Refinement. *Laboratory Animal 45*: 1-13. http://dx.doi.org/10.1258/la.2010.010031

Home Office 1995 *Code of practice for the housing and care of animals in designated breeding and supplying establishments*. The Stationery Office: London, UK

Honess P and Wolfensohn S 2010 A matrix for the assessment of welfare and cumulative suffering in experimental animals. *Alternatives to Laboratory Animals 38*: 205-212

Honess PE, Gimpel JL, Wolfensohn SE and Mason GJ 2005 Alopecia scoring: the quantitative assessment of hair loss in captive macaques. *Alternatives to Laboratory Animals 33*: 193-206

Lloyd MH, Foden BW and Wolfensohn SE 2008 Refinement: promoting the 3Rs in practice. *Laboratory Animals 42*: 284-293. http://dx.doi.org/10.1258/la.2007.007045

Morton DB and Griffiths PHM 1985 Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *Veterinary Record 116*: 431-436. http://dx.doi.org/10.1136/vr.116.16.431

NC3Rs 2007 National Centre for the 3Rs. *Guidelines on Primate Accommodation, Care and Use*. www.nc3rs.org.uk/page.asp?id=277

Sharpe SA, Eschelbach E, Basaraba RJ, Gleeson F, Hall GA, McIntyre A, Williams A, Kraft SL, Clark S, Gooch K, Hatch G, Orme IM, Marsh PD and Dennis MJ 2009 Determination of lesion volume by MRI and stereology in a macaque model of tuberculosis. *Tuberculosis 89*: 405-416. http://dx.doi.org/10.1016/j.tube.2009.09.002

Sharpe SA, McShane H, Dennis MJ, Basaraba R, Gleeson JF, Hall GA, McIntyre A, Clark S, Gooch K, Beveridge NER, Nuth E, White A, Marriott A, Dowall S, Hill AVS, Williams A and Marsh PD 2010 Establishment of an aerosol challenge model of tuberculosis in rhesus macaques and an evaluation of endpoints for vaccine testing. *Clinical and Vaccine Immunology 17*: 1170-1182. http://dx.doi.org/10.1128/CVI.00079-10

WHO 2014 *World Health Organization Tuberculosis Fact sheet No 104 March 2014*. http://www.who.int/mediacentre/factsheets/fs104/en/

Wolfensohn S and Honess P 2005 *Handbook of Primate Husbandry and Welfare*. Blackwell Publishing: London, UK. http://dx.doi.org/10.1002/9780470752951

Wolfensohn S and Lloyd M 2013 *Handbook of Laboratory Animal and Welfare*. Wiley-Blackwell: London, UK

https://doi.org/10.7120/09627286.24.2.139 Published online by Cambridge University Press