Netherlands National Committee
for the protection of animals
used for scientific purposes

Framework for Prioritising Animal Experiments for Replacement

Advice of the Netherlands National Committee for the protection of animals used for scientific purposes (NCad)

For the laboratory animals of today and the innovations of tomorrow

Netherlands National Committee for the protection of animals used for scientific purposes

NCad and its methods

The Netherlands National Committee for the protection of animals used for scientific purposes (NCad) is an independent advisory body dedicated to the protection of laboratory animal welfare. The committee fulfils its role by providing solicited and unsolicited advice, encouraging innovation and knowledge development, and bringing stakeholders together, so achieving tangible improvements focused on the replacement, reduction and refinement (3Rs) of animal experiments and the development of non-animal technologies.

Members of NCad
Top, from left to right: Henk Smid (chair), Reineke Hameleers, Coenraad Hendriksen.
Below, from left to right: Wim de Leeuw, Jan-Bas Prins, Roger Adan.

Cover photo
pexels.com

Contents

Introduction

It is currently still not possible to end the widespread use of animals for safety and toxicity tests and fundamental and translational research and instead use only non-animal alternatives. However, more and more New Approach Methodologies (NAMs)[1] are being developed.

In 2021, the Netherlands National Committee for the protection of animals used for scientific purposes (NCad) decided to investigate whether it might be possible to prioritise which animal experiments should be replaced with NAMs first. NCad was particularly concerned with animal models with limited translational value (while the translation of the research results really matters), animal models that involve severe distress, and animal models with little relevance or ethically questionable goals. NCad did not want to produce a simple list of candidates for replacement, but rather a number of criteria for prioritising animal models.

European Directive 2010/63/EU (Recital 12)[2] is based on the premise that animals do not represent only an instrumental and/or economic value. Animals should therefore always be treated as sentient beings. Their use in animal experiments should be limited only to areas of research that may ultimately benefit human, animal or environmental health. The starting point of the Animals Act is the intrinsic value of an animal.[3] Intrinsic value also forms the basis of the Experiments on Animals Act (Wod), in which the protection of the laboratory animal is paramount.[4]

The EU Directive states that it is desirable – and a final goal – to replace the use of animals in experiments with other methods that do not use live animals (recital 10, 2010/63/EU, note 2). The Directive recognises that this is currently often impossible. When choosing a method

[1] New Approach Methodologies (NAMs) are technologies and approaches that have the potential to provide the same or better safety and risk assessment information without conducting animal experiments. These include in silico, in chemico, in vitro and ex-vivo approaches, as well as the use of alternative species that are less perceptive (sentient) or of which fewer individuals are required. *The concept is now also used in biomedical research.*

[2] *Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes.* European Commission. Official Journal of the European Union 2010 Vol. L276 Issue L276 Pages 33-79.

[3] Preamble to the Animals Act: Whereas we have taken into consideration that, in order to implement the European obligations and in the interest of animal health and welfare and public health, it is desirable to lay down rules concerning animals, in particular animals kept by humans, recognising the intrinsic value of the animal (...). https://wetten.overheid.nl/BWBR0030250/2022-12-22 wetten.nl - Regulation - Animals Act - BWBR0030250 (overheid.nl).

[4] Preamble to the Experiments on Animals Act: Whereas we have taken into consideration that, with a view to the protection of animals, it is desirable to lay down rules with regard to the performance of experiments on animals. https://wetten.overheid.nl/BWBR0003081/2021-07-01.

involving the use of animals, the principles of replacement, reduction and refinement should be practiced, with the use of *alternative methods* explicitly having priority, if these methods are at least as suitable for answering a research question.

Once an alternative method has been recognised for regulatory research, it must also be applied according to the EU directive. This applies both to alternative methods that *no longer use laboratory animals* and to methods that *reduce* or *refine* the use of animals (recital 10, 2010/63/EU, see note 2). The latter methods are subject to the following guidelines (recital 11, 2010/63/EU, see note 2):

- Choose the method that produces the best results and is likely to cause the least pain, suffering and anxiety.
- Use the smallest number of animals that gives reliable results.
- Use animal species that are least susceptible to pain, suffering, distress or lasting harm and most suitable for extrapolating results to the target species.
- As far as possible, avoid death as an endpoint of an animal experiment and replace it with early-stage human endpoints (or better: human intervention points[5]). If death as a scientific endpoint is inevitable, do whatever is possible to limit suffering.

This means that if it is not possible to replace an animal model with a non-animal alternative, it is desirable to consider whether the animal experiment can be replaced by another animal experiment with less impact on the animals, or by using fewer animals, or by minimising animal distress.

When granting a licence for animal experiments, a committee – in the Netherlands this is the Central Authority for Scientific Procedures on Animals, advised by local animal ethics committees – must assess the application against a number of aspects, including (Article 18, 2010/63/EU, see note 2):

- the objectives of the project and the predicted scientific or societal benefits or educational value;
- the attention paid to replacement, reduction and refinement;
- the assessment of the expected maximum distress.

A *harm-benefit analysis* is then conducted to assess whether the suffering or distress is justified by the expected benefit to humans, animals or the environment, taking into account ethical considerations (Article 38(2), 2010/63/EU, see note 2).

[5] W. O. Williams and P. Baneux. *Humane Intervention Points: Refining endpoint terminology to incorporate non-euthanasia intervention options to improve animal welfare and preserve experimental outcomes.* Lab Anim 2022 Vol. 56 Issue 5 Pages 482-489. https://doi.org/10.1177/00236772221090801.

In light of the above, NCad asked Corné Rademaker, ethicist and member of several animal ethics committees, to write a report on prioritising animal experiments for replacement and investigate criteria that could be used to assess which animal models should be prioritised for the development of NAMs. The report prepared by Rademaker is a comprehensive literature review of criteria by which animal models could be prioritised. The report *Prioritising Animal Experiments for Replacement* is attached as Annex 3.

The following is an initial, extensive summary of Rademaker's report. This is followed by NCad's advice in response to this report. We also consider the implications of the criteria suggested by Rademaker for the ethical review framework. The final section discusses how the criteria could be used to prioritise the replacement of animal models. We end this advisory report with a discussion of the possible follow-up to this study.

1.

Prioritising animal experiments for replacement - summary

In the report *Prioritising Animal Experiments for Replacement* (Annex 3), Rademaker discusses criteria against which an animal experiment could be assessed and which can help to decide whether such an experiment should have priority for phasing out. His thorough literature review examines the grounds for such a prioritisation. He discusses eight possible criteria and concludes that four of them are useful for the stated purpose. Rademaker rejects four criteria:

- The degree of proportionality is a measure of the suitability, necessity and adequacy of an animal experiment in the context of reducing harm, and itself refers to other criteria;
- The stage of development of animal free/non-animal alternatives as potential 'low-hanging fruit', but this does not necessarily involve research that requires phasing out on ethical grounds. Of course, solutions in an advanced stage of development still deserve attention[6];
- The number of animals (animal experiments) is an important factor in relation to other characteristics, such as the goal of the experiment or the distress caused to the animals, but is less suitable as a criterion in itself[7];
- The species of animal used may be relevant in relation to the validity of the animal experiment, and therefore the translatability and the ability to experience distress. Both these criteria are included in Rademaker's report as independent criteria. There is no consensus among philosophers and ethicists on whether some species are fundamentally more worthy of protection than others. However, the EU Directive prohibits the use of a number of primate species (Article 8, 2010/63/EU, see note 2) and sets stricter rules for other primate species than for other species of animals[8].

Rademaker then elaborates on what he believes to be the four suitable criteria, namely the degree of translatability, the relevance of the goal of the experiment, the degree of distress, and the degree of integrity violation.

[6] NCad's response in Chapter 3: 'NCad's reflection on the report'.
[7] NCad's response in Chapter 3: 'NCad's reflection on the report'.
[8] NCad's response in Chapter 3: 'NCad's reflection on the report'.

Translatability

Translatability concerns the extent to which the data obtained from an animal experiment are translatable to humans or another target species. Translatability also concerns validity: an animal model has validity only insofar as it is a reliable model of the disease, disorder or phenomenon it represents, and more specifically insofar as the same underlying mechanisms are involved.

The translatability of animal experiments has long been an issue in biomedical research. A recent systematic review (Leenaars et al.[9]) reveals that the predictive value of findings of experimental animals for phase I/II *clinical trials* can vary enormously.

Rademaker's report discusses various aspects of validity in detail. At least three aspects of validity are distinguished in the literature:

- Predictive validity concerns the ability to predict a human or animal phenomenon based on the performance of an (animal) model.
- Face validity refers to the degree of similarity between the (animal) model, and the disease or condition under study, across the widest possible range of symptoms.
- Construct validity concerns the extent to which the mechanisms underlying the disease or condition correspond with the (animal) model and the target species[10].

Then there is internal validity (reliability and reproducibility) and external validity (broad applicability; generalisability), which are more determined by a combination of the characteristics of the model and the experimental context in which the model is used. *Reproducibility* indicates whether the model produces consistent and reliable results. *Generalisability* concerns the question of whether the results obtained with the animal model in the population or breeding line used can be extrapolated to the target organism.

In addition, a distinction is made between convergent validity and distinctive validity. Convergent validity indicates that interventions that measure the same thing should correlate significantly with each other. Distinctive validity is exactly the opposite.

The validity criteria must be applied to research where translatability is a goal, such as all regulatory and preclinical translational research. Fundamental research is mainly concerned with unravelling biological mechanisms and the follow-up research questions about the extent to which the mechanism found is more or less generally applicable. The results of fundamental research contribute to answering the validity question, i.e. to what extent the subject of a study can be modelled by, for example, a specific cell, or organ system, or animal model. Rademaker says some research classified as fundamental research is actually also translational in nature.

He goes on to question how translatability could be used as a criterion.

According to Rademaker, the translatability of an animal model depends on the degree to which one can state, with certainty, that any causal relationship found cannot be explained by factors other than the factor that was the goal of the research (internal validity), and on the broad applicability of the results (external validity). Both validities say something about the extent to which the model, and the way it was used in the experiment, simulates the presumed underlying mechanisms of a biological or etiological phenomenon in humans or the target species.

Rademaker also discusses four instruments that can be used to establish the degree of translatability. Several researchers have developed scoring systems to evaluate the validity of animal models. Rademaker recommends selecting one or more of these scoring systems in consultation with the professional field.

Finally, Rademaker asks whether translating research results between one species and another is even possible. There are examples of both successes and failures. Animal species, and breeding lines within an animal species, clearly differ from each other. The question is whether this also means that an animal can only serve as a model for animals of the same species (or breeding line), and whether an animal model can at most offer a hypothesis about another species. In the latter case, a model will be revealed to be translatable *only after* the results have been compared with the results in the target animal or human.

Rademaker also raises the question of whether this conclusion means that developing new, unproven animal models is therefore ethically more problematic than when using proven models. His answer is that it is not always necessary for the causal mechanism to be exactly the same. Moreover, hypothesis-driven models also have value.

Animal experiments that use animal models with a low degree of translatability have first priority for replacement, also depending on the assessments of other criteria, such as the relevance of the goal of an animal experiment.

---

[9] Leenaars, Cathalijn H. C., Carien Kouwenaar, Frans R. Stafleu, André Bleich, Merel Ritskes-Hoitinga, Rob B. M. De Vries, and Franck L. B. Meijboom. *Animal to Human Translation: A Systematic Scoping Review of Reported Concordance Rates.* Journal of Translational Medicine 2019 Vol. 17 Issue 1 Pages 1-22. https://doi.org/10.1186/s12967-019-1976-2.

[10] For example, the target species will be humans if the animal model is used to study a human disease.

The relevance of the goal

Animal experiments are conducted with a goal. The immediate goal of a concrete animal experiment (e.g. unravelling a mechanism) can be distinguished from the ultimate goal which lies further in the future and requires multiple experiments and studies.

Some of those ultimate goals are the subject of debate. For instance, many concerned parties ask if we cannot find an alternative to routine safety testing of chemicals on animals. In contrast, research focused on improving human and animal health scores high as a goal. Even within those categories, some goals are more justifiable than others, as some conditions can also be addressed or made more bearable by other means. Medical research is often focused on the cure, while prevention can also be an option (such as nutritional interventions to strengthen the immune system). In research for the benefit of the livestock industry, it matters whether such research focuses on animal health and welfare, or developing a more sustainable livestock industry, rather than only on increasing production (Council on Animal Affairs, 2018).

There is no simple hierarchy that can be established to this end. Rademaker thinks it would be helpful to distinguish between basic, serious and peripheral interests (VanDeVeer[11]):

- Basic interests involve specific conditions or goods that are required by humans and animals to enable "minimally adequate" functioning. Rademaker applies this distinction to the relevance of research goals. Research with a basic goal could involve studies aimed at preventing, curing and making more tolerable life-threatening diseases and diseases that seriously impair the quality of life of humans and animals, and the conservation of animal species and ecosystems.
- Serious interests involve goods or conditions that are important for human and animal welfare. Serious goals could include studies aimed at developing sustainable forms of livestock farming and food safety tests.
- Peripheral interests involve goods and conditions aimed at making life pleasant. Examples include decorations, certain grooming items, or fur garments. In today's modern society, livestock research to improve production could be called a peripheral goal.

These can all be legitimate goals of research, but a hierarchy of importance can be established when the use of laboratory animals is required to achieve such a goal.

[11] VanDeVeer, Donald. (1995) *Interspecific Justice*. In: *People, Penguins, and Plastic Trees: Basic Issues in Environmental Ethics*, edited by Christine Pierce and Donald VanDeVeer, 85-99. Belmont, CA: Wadsworth Publishing Company.

Distress

Research involving animals is only called an animal experiment if it involves a certain minimum threshold level of distress. European Directive 2010/63/EU (see note 2) makes the following distinction, with two thresholds:

- Below the distress threshold; the distress is lower than the distress caused by inserting a hypodermic needle according to good veterinary practice. Such research is not considered to be an animal experiment.
- Terminal; the animal is anaesthetised; all interventions are performed under anaesthesia; the animal does not recover.
- Mild distress;
- Moderate distress;
- Severe distress;
- Very severe distress (i.e. severe pain, suffering or fear that is likely to be long lasting and cannot be alleviated) is allowed only in very exceptional cases and with the explicit authorisation of the minister, and falls outside normal procedures. No animal experiments involving 'very severe distress' have been reported in the Netherlands or Europe.

In prioritising the replacement of animal experiments on the basis of distress, Rademaker suggests that we should focus first on animal experiments with animal models that cause severe distress. This concerns about 1% of all animal experiments in the Netherlands[12] and about 11% in Europe[13]. Within the EU (incl. Norway), *batch potency testing* of vaccines and other biological products is the main source of severe distress, followed at some distance by nervous system research and diagnostic tests.

Integrity violation

The debate about genetic modification arose in the Netherlands in the early 1990s in response to the advent of genetically modified (GM) animals such as the bull Herman. People apparently experience moral unease when confronted with GM animals, even if no violation of health or welfare is involved. A term was subsequently sought to describe this phenomenon. Initially, the terms 'wholeness' and 'naturalness' were used. Later the term 'integrity' was adopted. This term was also suitable for describing other 'violations' of animals that are not directly related to welfare or health. Besides genetic modifications, it is also applied to animals that are bred for certain traits or who have their beaks, horns or reproductive organs removed.

[12] p. 25 https://www.nvwa.nl/documenten/dier/dierenwelzijn/zo-doende/publicaties/zo-doende-2020-jaaroverzicht-dierproeven-en-proefdieren.

[13] p. 32 https://ec.europa.eu/environment/chemicals/lab_animals/pdf/SWD2019_Part_A_and_B.pdf.

According to the definition of Rutgers and Heeger,[14] integrity concerns the wholeness of the animal, the species-specific balance of the animal, and the ability of the animal to sustain itself independently in an environment appropriate for the species.

Rutgers and Heeger cite the dehorning of cattle as an example of the violation of the *wholeness* of the animal. For the *violation of an animal's species-specific balance*, Rutgers and Heeger give the example of breeding for rapid muscle growth in broilers, which eventually restricts the animal's capacity to walk and sometimes even results in spontaneous death well before the animal is ready for slaughter. This is because the metabolism is out of balance in these broilers. An example of a violation of the animal's *capacity to sustain itself independently in an environment appropriate for the species* is the breeding of 'double-muscled' cattle like the Belgian Blue. These animals show exceptionally strong muscle growth but have a narrow birth canal, so the calves of double-muscled breeds cannot be born naturally way, but must always be delivered by a caesarean section.

Rademaker then discusses the response to Rutgers and Heeger's definition of integrity by later authors. He sees that two interpretations of *wholeness* are applied, namely:

- Interventions in the animal body. Amputation (e.g. toe clipping) falls under this interpretation, while a breeding result (e.g. nude (furless) mice) does not. This is a narrow interpretation of wholeness.
- Lacking something that an animal naturally, as a species, physically possesses (e.g. fur, horns). This is the broad interpretation of integrity and also includes a breeding result. The physical configuration of an animal gives it all kinds of abilities. Some animals are born naturally without fur and a breeding line is then created with these animals (hairless cats, nude mice). These animals have trouble maintaining their body temperature and need an artificial environment to survive.

The broad interpretation of wholeness includes the capacities or *abilities* of animals in the interpretation of integrity. This is also in keeping with the original definition of Rutgers and Heeger, which speaks of the *ability* of the animal to sustain itself independently in an environment appropriate for the species.

Rademaker notes that this focus on animal abilities has only become stronger over time. However, Rademaker criticises the strong focus on independence in the definition of integrity. He prefers the concept of *species-specific abilities* of animals, such as gnawing behaviour in rodents. This means that animals should have the opportunity to exhibit species-specific behaviour.

According to Rademaker, the broad interpretation of integrity does the most justice to people's sense of moral unease concerning, say, nude mice or double-muscled cattle. Rademaker then addresses some of the questions raised and discussed by philosophers in the debate that was conducted during the 1990s.

Can we only speak of an integrity violation if the violation is for utilitarian purposes?
In other words, does the intention matter when, for example, a dog's tail is docked? This could indicate that integrity is not a trait or biological characteristic of an animal (such as welfare) but rather an interpretation of people's moral unease when confronted with certain interventions in a certain context. However, Rademaker argues that, even if an intervention is for the benefit of the animal itself, it can still involve an integrity violation. The amputation of a tail violates the wholeness of the animal, but may be necessary for its health. The principle of benevolence is more important here. Rademaker concludes that, like the concept of welfare, integrity is an evaluative concept that also has an empirical side which can be determined based on the consequences of an intervention for the animal.

Is integrity an absolute or a relative concept?
Integrity seems to be an absolute concept: the integrity of the animal is violated or it is not. However, Rademaker raises the question of whether there are also *degrees* of integrity violation (similar to welfare). Referring to Brom,[15] he argues that this may indeed be the case. An animal whose integrity has been violated (e.g. its tail docked) may have its integrity violated again at a later date (its ears are docked too).

Does integrity only concern individual animals, or does it also include breeding lines or species?
This question mainly arises in the field of genetic modification (GM). GM takes place at the early embryonic stage and does not necessarily affect the phenotype. So the question is, when does one speak of an integrity violation? Presumably, this will not be the case as long as no phenotypic change is identified. If altered genetic traits are passed on to offspring, those traits are subsequently innate to the offspring. Has the integrity of these offspring been violated? It depends on what you take as your reference point: the wild-type animal, the breeding line (or breed), or the observable characteristics or traits of the animal (the phenotype). If the breeding line is taken as a reference point, changes in the current generation do not count; they are taken for granted. That does not do anything to alleviate the moral unease people

[14] Rutgers, B. and Heeger, R. (1999) *Inherent worth and respect for animal integrity*. In: *Recognizing the intrinsic value of animals: beyond animal welfare*, Van Gorcum, Assen, The Netherlands.

[15] Brom, Frans W.A. (1997). *Onherstelbaar verbeterd: Biotechnologie bij dieren als een moreel probleem*. Assen: Van Gorcum.

may experience with such a breeding result. The other two cases may well be cases of 'breeding involving an integrity violation', analogous with 'breeding involving a welfare violation'.

Is integrity a usable criterion in professional practice?
To use integrity violation as an assessment criterion, we must be able to assess the violation, as we would do with a welfare violation (distress). This leads to the question of whether there are degrees of integrity violations. In the narrow definition (physical wholeness), only the number of violations counts (e.g. a vasectomy as well as a toe clipping). In the broad definition, the number of violations also plays a role, but in addition, any violation of species-specific abilities is also counted as a criterion. Species-specific abilities may be affected to a greater or lesser extent. As with animal welfare, this in any case concerns developing an intersubjectively quantified assessment.

However, it is not easy to operationalise the broad interpretation of integrity, because this requires a lot of knowledge of the species-specific abilities of animals and the violation thereof.

Because no instruments have yet been developed to assess integrity violations, Rademaker recommends establishing a working group to work out a proposal for a classification of integrity violations. This could take the form of a classification in terms of mild, moderate and severe degrees of integrity violation, linked to the frequency of interventions, the degree to which a specific animal ability is affected, and the number of abilities affected.

Conclusion

Rademaker's brief was to develop criteria to prioritise animal models for replacement with alternatives. Rademaker suggests applying four criteria and giving priority to the replacement of animal models that:

- Have a low degree of translatability and/or
- Serve peripheral goals and/or
- Cause severe distress and/or
- Seriously violate the integrity (species-specific ability) of animals

2.

NCad's advice

The Rademaker report is a clear exposition of criteria based on which it is desirable to phase out certain animal models and, where possible, replace them with alternatives in the broad sense (the so-called NAMs, see note 1). In particular, the elaboration of the concepts of translatability and integrity violation have enriched the debate on the acceptability of specific animal models. They also help to delineate the harm-benefit analysis.

Rademaker discusses eight criteria, and concludes that four of these are suitable for prioritising animal models that should be rejected or replaced. He goes on to discuss each criterion separately. In practice, a harm-benefit trade-off is always made and the criteria interfere with each other. Sometimes, a model that you would want to reject on the basis of one criterion has to be retained on the basis of another. This could be because there is no non-animal alternative available yet and it does not look like one will be developed in the foreseeable future, while the experiment is considered very important. For these cases, NCad stresses the importance of following the Rs of Reduction and Refinement.

In its advice, NCad reflects on Rademaker's report (Annex 3), conducts a brief additional discussion of the integrity criterion, and also further defines the criterion of translatability. It then discusses what these criteria mean for the harm-benefit analysis (the ethical assessment) and suggests how the criteria developed by Rademaker could contribute to the harm-benefit analysis (as carried out by animal welfare bodies, animal experiment committees and the Central Authority for Scientific Procedures on Animals), as well as to the discussion around this harm-benefit analysis. The final section is aimed at policymakers and funding bodies and provides advice on how they can prioritise animal models for replacement and help steer the development of NAMs. Finally, NCad suggests a number of follow-up steps.

The NCad's reflection on the report

Animal species used

Rademaker argues that distinguishing between animal species is difficult on ethical grounds and that there is no consensus among philosophers on this issue. He therefore does not elaborate on using animal species as a criterion. Currently applicable legislation is generally understood to reflect the norms and values of a society.[16] The EU Directive (Article 8, 2010/63/EU, see note 2) and the Dutch Experiments on Animals Act (Articles 1(5) and 10e, Experiments on Animals Act, see note 4) *do* distinguish between animal species. An experiment on an invertebrate animal is not an animal experiment, cephalopods excepted. So, experiments involving invertebrates (cephalopods excluded) do not require a project licence and are not regulated. Primates have been granted a special status in the EU Directive and the Dutch Experiments on Animals Act, with stricter requirements for their use for scientific goals. In line with the EU Directive and the Experiments on Animals Act, replacing animal models with non-human primates would therefore have priority. This distinction between animal species is the legal reflection of the notion that, the more similar a species is to humans, the more objectionable the use of that species.

Rademaker argues that this aspect of the European Directive can be questioned based on ethical considerations, and the NCad agrees. NCad also believes that the difference in sensitivity of, and/or the degree of distress experienced by, an animal (based on objective criteria) is a more important criterion than the type of animal species. A particular intervention may have more impact on a specific species, or within a species on a specific breeding line, than on another. If there are no objective criteria, the precautionary principle applies.[17]

The stage of development of alternatives

Rademaker considers the development stage of a non-animal alternative to a specific animal model to be of strategic importance, but does not see that development stage as an ethical consideration for prioritising an animal model for phasing out. His report focuses on the animal models that are most problematic, not the 'low-hanging fruit'. NCad agrees in part. Every model has useful and less useful characteristics and/or limitations. It is up to science to recognise these and strive to continuously improve the models being used, and to develop new, better models. This involves making choices and prioritising resources, including those deployed for the development and validation of NAMs. In this light, the development stage and availability of a NAM does form a criterion, all the more if it involves an alternative to a type of animal experiment that requires large numbers of animals (see below).

Ultimately, the licensing authority (the Central Authority for Scientific Procedures on Animals) decides whether or not to allow an animal experiment, and under what conditions. The applicant must provide convincing evidence that the proposed animal or animal model can reasonably be used to find the answer to the research question, and that there is no non-animal alternative available, or a model that requires fewer animals and/or involves less distress, which can achieve the same scientific objectives. A description of the useful and less useful characteristics of the various models, and the efforts being made to develop models further (including NAMs), should therefore be part of the application. If a research team plans to use animals for fundamental research, it will ideally clarify how the research results are expected to contribute to future – and different stages of – applied research (see below under 'Translatability').

Numbers of animals or animal models

NCad disagrees with Rademaker's rejection of the criterion for numbers of animals and animal models. The Dutch Experiments on Animals Act and the underlying EU Directive do not refer to the 3Rs (replace, refine and reduce) for nothing. 'Number as a criterion' does not so much refer to the number of animals per individual project licence application, where the minimum number of animals must be justified (the R for Reduction), but rather to the cumulative number of animals required for an animal model that is frequently and widely used in certain areas of research. This includes animal models that are widely used, year after year, to test the safety and efficacy of chemicals, drugs and vaccines. Changes in the regulations on safety could bring about the required change here. NCad therefore proposes including the number of animals used as a criterion for prioritising animal models for replacement (in the broad sense), or for refinement.

The concept of translatability of research results in fundamental research

The relevance of the translatability criterion for translating the research results of an animal model to, for example, human beings (in research labelled as translational) is easier to justify than fundamental research, which is mainly concerned with unravelling biological mechanisms, and the follow-up research questions about the extent to which the mechanism found

[16] Note: Norms and values are always in flux, albeit gradually, and the legislation follows them at some distance. That there is a debate among ethicists may indicate that such a change may be underway in society, where animals are seen as increasingly equal to each other and to humans.

[17] The precautionary principle is a moral and political principle that states that, if an intervention or policy measure may cause serious or irreversible damage to society or the environment, the burden of proof lies with the proponents of the intervention or measure, if no scientific consensus exists on the potential future harm.

is more or less generally applicable and/or has translational value.[18] The European Commission's 'Working document on Project Evaluation and Retrospective Assessment' provides some guidance on how to interpret harm-benefit analyses in the context of ethical reviews of project licence applications (see note 18). It considers the advancement of knowledge as sufficient justification for carrying out that research, provided it is sufficiently scientifically substantiated and the results are disseminated (taking into account intellectual property rights), as well as the possible longer-term benefits, whereby translatability becomes increasingly important. These aspects must be addressed when formulating the goal of the project, which must also describe the current state of knowledge on which the project aims to build (see also Annex 4). But tThis does not change the fact that the fundamental research in this field should also focus on the development and deployment of NAMs.

The harm-benefit analysis and Rademaker's criteria

The harm-benefit analysis is primarily concerned with weighing the relevance of an animal experiment against the distress caused to the laboratory animals. This assessment is carried out by the CCD following an advice submitted by an animal experiment committee in response to a project licence application.[19] The project licence application is prepared by the applicant researcher based on the advice of the Animal Welfare Body. This working method ensures that the quality of licence applications is high and animal welfare aspects are guaranteed.

The European Commission's 'Working document on Project Evaluation and Retrospective Assessment' (see note 18), which is used as a guidance by the CCD and the animal experiment committees, recommends using the modified Bateson's cube (Annex 2 of the Working Document) for the assessment framework (Figure 1). This cube has three axes: one for the relevance of the goal (the benefits), one for the distress (the harm), and a third for the probability of the benefits actually being achieved. The opportunities for translatability contribute to this probability and are therefore included in the harm-benefit analysis in practice.

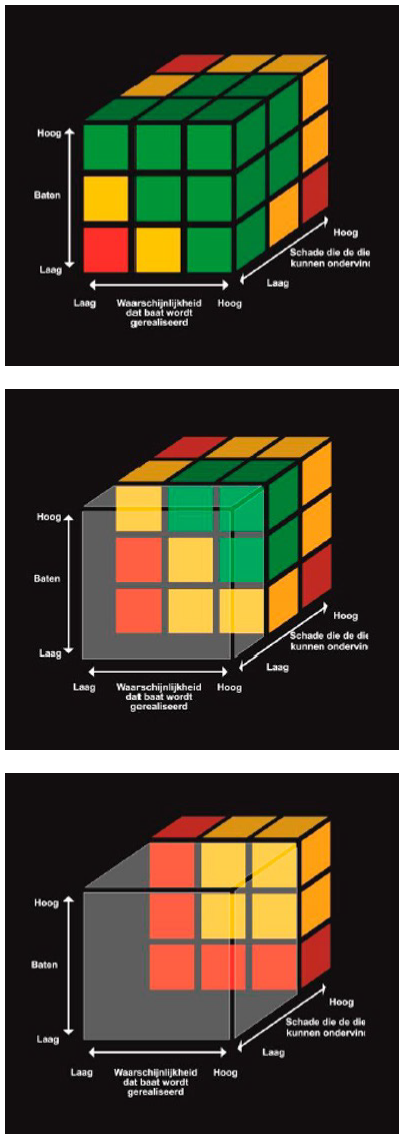
In his report, Rademaker suggests that integrity violation should also from part of the ethical assessment. Including integrity violation in the harm-benefit analysis could enrich how 'harm' is defined and help to refine the assessment framework, which is in line with the intention of the legislator. The CCD already applies this enriched definition of harm, as evidenced by its own Ethical Review Framework (see note 19).

Rademaker also notes that the concept of integrity violation is insufficiently operationalised. He mentions the concept of 'violation of a species-specific ability' as a possible starting point for this conversation. No instruments are yet available that could be used to distinguish between mild, moderate or severe integrity violations. The debate that played out among ethicists and philosophers in the 1990s does not offer guidance here either. Rademaker recommends making a start on this operationalisation. As was done in the past for welfare violations, we could work on reaching intersubjective agreement on what exactly is meant by a mild, moderate or severe integrity violation. To arrive at such a classification, we will first need to discuss how we think it should work. So, in case of potential integrity violations, we should try to describe what those integrity violations entail and consider how this should be factored into the ethical assessment. As recommended by Rademaker, a working group could consider concrete cases from professional practice.

[18] Annex 4 reproduces the texts on fundamental research in the *Working document on Project Evaluation and Retrospective Assessment* that was compiled by the working group of National Authorities responsible for implementing Directive 2010/63/EU.

[19] https://www.centralecommissiedierproeven.nl/binaries/ccd/documenten/formulieren/16/6/6/ethisch-toetsing-skader/praktische-handreiking-ethisch-toetsingskader-versie-september-2019.pdf.

Figure 1: Bateson's cube as modified by the EU working group (see note 18)



Prioritising the development of NAMs

The research question is central to any research project. The research strategy to answer that question requires careful selection of a model or combination of models to be used, including NAMs (see note 1). The four criteria described in Rademaker's report point to a direction for prioritising the development of NAMs (in the broad definition given earlier). NCad adds to Rademaker's four criteria the 'number of animals' and the development stage of non-animal alternatives. However, a commitment to the other two Rs of reduction and refinement will remain important as long as animal experiments are necessary. To this end, NCad has established a separate project for the prevention of severe distress. The RSPCA's (vision) document "A 'Road Map' Toward Ending Severe Suffering of Animals Used in Research and Testing"[20] has been translated into Dutch. NCad is organising a symposium on this topic in autumn 2023, together with the Platform of Animal Welfare Bodies, the RSPCA and Leiden's Animal Welfare Body.

Subsequent questions then concern whether these criteria can be prioritised, how the criteria might reinforce each other, and what those criteria mean for the research and the spending of resources on the development of alternatives.

We have seen above that it is worthwhile to at least take a closer look at the harm-benefit analysis. If animal experiments have little relevance because they serve only a 'peripheral' interest, or because the knowledge generated by the animal experiment is of little relevance to the stated goal (for example because the animal model has a low degree of translatability), then it is unlikely that such an experiment will be approved. To answer their research question, the researchers would need to find a non-animal alternative, or an animal model with very little impact on animals, and if such models are not available, they would have to be developed.

This is different where 'basic' or 'serious' interests are involved, and the trade-off between benefits and harm (distress and integrity violation) will play a more prominent role. If the distress is severe, or if the integrity of the animals is severely violated, the development of alternatives in the broad sense, including refined models, should certainly have priority.

Priority should also be given to developing NAMs to replace commonly used animal models that – due to their frequent use – require large numbers of animals over time. Applications for projects involving these – often prescribed – models do not need to be rejected outright, but rather it is desirable that these models are no longer used.

[20] Lilley, E., Hawkins, P. and Jennings, M. (2014) *A 'road map' toward ending severe suffering of animals used in research and testing.* Altern Lab Anim 2014 Vol. 42 Issue 4 Pages 267-72.

Finally, priority should also be given to further developing and validating NAMs currently under development, particularly if they form an alternative to commonly used models.

Conclusions and recommendations

- Prioritisation is a useful instrument for ensuring the optimal deployment of resources for the development and implementation of NAMs and the replacement of animal models. Rademaker's report and this NCad advisory report can contribute to achieving this. NCad will bring both documents to the attention of the key funding bodies: NWO, ZonMw, the health funds and ProefdierVrij.
- NCad asks the Minister of Agriculture, Nature and Food Quality to inform the Minister of Education, Culture and Science of this advisory report, and suggests that additional resources be earmarked for the development and implementation of NAMs together with the other ministries. The Minister of Education, Culture and Science can also be asked to enter into discussions with NWO and ZonMw, based on this advisory report and the Rademaker report, to determine how these funds could be prioritised.
- It is recommended to elaborate on, and include the concept of translatability in, the 'probability' axis of the modified Bateson's cube. To this end, NCad advises the Minister of Agriculture, Nature and Food Quality to promote including translatability in the 'Working document on Project Evaluation and Retrospective Assessment' (see note 18) and later also in the evaluation of Directive 2010/63/EU (see note 2).
- It is recommended to elaborate on, and include the concept of integrity violation in, the 'harm' axis of the modified Bateson's cube. This will serve to enrich the harm-benefit analysis. To this end, NCad advises the Minister of Agriculture, Nature and Food Quality to promote including integrity violation in the 'Working document on Project Evaluation and Retrospective Assessment' (see note 18) and later also in the evaluation of Directive 2010/63/EU (see note 2).
- The distinction made between animal species in Directive 2010/63/EU is questionable. NCad believes that it is not the relationship with, or similarity to, humans, but rather the degree of distress experienced by the animal (its sensitivity), based on objective criteria, which is relevant when choosing an animal species or breeding line. A particular intervention may have more impact on a specific species, or within a species on a specific breeding line, than on another. NCad advises the Minister of Agriculture, Nature and Food Quality to share these considerations with the European Commission during the evaluation of Directive 2010/63/EU.
- Safety regulations require animals to be used to test the safety and efficacy of chemicals, drugs and vaccines, year after year, and at a huge scale. NCad advises the minister of Agriculture, Nature and Food Quality to endeavour to have the international safety regulations adapted to enable reduction of the use of animals for testing the safety and efficacy of substances.

Annex 1 Authors

NCad working group members

Dr Monique Janssens (until July 2022)
Prof. Jan-Bas Prins (from July 2022)
Dr Elmar Theune (NCad staff)

External expert

Dr Corné Rademaker - author of the report 'Prioritising Animal Experiments for Replacement'.

Annex 2
List of abbreviations

3Rs	Replacement, Reduction and Refinement
CCD	Central Authority for Scientific Procedures on Animals (*Centrale Commissie Dierproeven*)
DEC	Animal Ethics Committee (*Dierexperimentencommissie*)
EU	European Union
GM	enetically Modified
IvD	Animal Welfare Body (*Instantie voor Dierenwelzijn*)
LNV	Ministry of Agriculture, Nature and Food Quality
NAM	New Approach Methodology (aka: New Approach Method)
NCad	Netherlands National Committee for the protection of animals used for scientific purposes (*Nationaal Comité advies dierproevenbeleid*)
NWO	Dutch Research Council
OC&W	Ministry of Education, Culture and Science
RSPCA	Royal Society for the Prevention of Cruelty to Animals
Wod	Experiments on Animals Act (*Wet op de dierproeven*)
ZonMw	Dutch organisation for health research and innovation in healthcare (*Nederlandse organisatie voor gezondheidsonderzoek en zorginnovatie*)

Annex 3

Corné Rademaker's report entitled 'Prioritising Animal Experiments for Replacement'

Corné J. Rademaker

Version date: 17/03/2022

Summary

This report was commissioned by the Netherlands National Committee for the protection of animals used for scientific purposes (NCad). The aim of the report is to provide criteria that can be used to prioritise full or partial (relative) replacement of the current animal experiments. Four criteria were developed based on interviews and a literature review:

1. the degree of translatability of the research;
2. the degree of relevance of the goal of an animal experiment;
3. the degree of distress caused to laboratory animals; and
4. the degree of violation of the integrity of laboratory animals.

The first criterion, translatability, concerns determining the reproducibility and generalisability of animal models used in animal experiments. Animal experiments that use animal models with a low degree of translatability have first priority for replacement. The degree of translatability can be estimated based on the tools that have already been developed for this purpose. The second criterion, the relevance of an animal experiment, can be separated into basic, serious, and peripheral interests, with animal experiments that serve peripheral interests having the highest priority for replacement. The third criterion, the degree of distress caused to an animal, concerns the pain, suffering, fear or lasting harm that an individual animal may suffer during an animal experiment. Animal experiments that cause severe and prolonged distress have first priority for replacement, followed by any animal experiments involving severe distress. The last criterion, the degree of violation of integrity, refers to any physical intervention (narrow interpretation of integrity), or in fact any intervention that results in a loss of species-specific abilities (broad interpretation of integrity). Animal experiments that constitute a severe violation of integrity have the highest priority for replacement. However, the classification of the degree of violation of integrity should be further elaborated, for example by a working group (analogous to the classification of distress in animal experiments). Recommendations for stakeholders regarding animal experiments are presented at the end of the report. The report concludes that the aforementioned four criteria comprise a useful tool for identifying which animal experiments and animal models should be prioritised for the development of alternatives.

Contents

Introduction

The current policy in Europe and the Netherlands on animal experiments is based on the implementation of the 3Rs (Replacement, Reduction and Refinement) and a harm-benefit analysis. The 3Rs policy is based on the work of Russell and Burch (1959). Replacement occupies a special place in the 3Rs methodology. According to Sandøe et al. (2015), Russell and Burch formulated replacement as the first of the 3Rs for a reason: reduction and refinement are relevant only when there are no opportunities for replacement. Moreover, (full) replacement is the only 'R' that is fully in line with the ideals and objectives of animal rights organisations and action groups. Consequently, the replacement alternative has widespread support in society (cf. Smith and Boyd, 1991, 134-36).

In fact, European Directive 2010/63/EU regulating the use of animals in animal experiments describes complete replacement as its ultimate goal: "However, this Directive represents an important step towards achieving the final goal of full replacement of procedures on live animals for scientific and educational goals as soon as it is scientifically possible to do so" (European Parliament and Council of the European Union, 2010, 34). Recently, the European Parliament adopted a resolution calling for an acceleration of the transition to non-animal methods in order to achieve the ultimate goal of Directive 2010/63/EU (European Parliament, 2021).

In the Netherlands, the Experiments on Animals Act regulates the use of animals for "experimental or other goals, with known or unknown outcome, or for educational goals, which may cause the animal a level of pain, suffering, fear or lasting harm equivalent to, or higher than, that caused by the introduction of a needle in accordance with good veterinary practice" (Experiments on Animals Act, 2021, §1, Article 1a). The Experiments on Animals Act requires project proposals to be ethically reviewed by the Central Authority for Scientific Procedures on Animals (CCD), itself advised by animal ethics committees (DECs). The animal ethics committees and the Central Authority thereby review proposed animal experiments against a statutory framework, including options for replacing animal experiments. The challenge for all other parties, such as researchers, the industry, ministries, and the Netherlands National Committee for the protection of animals used for scientific purposes (NCad), is to continue to strive for, accelerate and promote the development and recognition of non-animal alternatives. To this end, NCad sees a need to prioritise the replacement of animal experiments. While the ultimate aim is to replace animal experiments completely, a strategy is needed to identify those animal experiments that pose the greatest ethical challenge, so that the efforts to replace them completely or relatively can be focussed.[21] Some animal experiments are more hotly debated than others in the social and scientific arenas, for various reasons. For example, Maastricht University ceased conducting heart failure research in Labradors in 2015 following protests against this type of testing (Animals Today, 2015; NOS, 2015), and the forced swim test used in depression research is increasingly under fire because of the distress and validity problems involved (PETA, 2022; Molendijk and De Kloet, 2019). Taking this social and scientific debate as its starting point, this study distils and elaborates criteria that may be helpful to prioritise which animal experiments should be replaced first. The premise is that animal experiments that are the least ethically acceptable should be replaced first, and the identified criteria should reflect this.

NCad's ultimate goal is to produce an advisory report that outlines criteria on which to base the prioritisation for replacing certain animal experiments. The objective of the present report is to make a first contribution to this ultimate goal based on a series of interviews and a literature review.

[21] Russell and Burch (1959, Ch. 5) distinguish between 'full' or 'absolute', and 'relative' replacement. Absolute replacement is the complete replacement of laboratory animals with non-sentient material, and constitutes the absolute ideal. Russell and Burch describe relative replacement as the use of animals "exposed, probably or certainly, to no distress at all" (Russell and Burch, 1959, Ch. 5 'Modes of Absolute and Relative Replacement'). Both options for replacement are implied in this report, whereby it is important to note that in addition to distress, the violation of integrity also plays a role. From the perspective of relative replacement, not only are animal experiments that cause less distress preferred, but also animal experiments that involve a lesser violation of integrity.

Methodology

In late 2021, stakeholders in the field of Dutch laboratory animal science were contacted to help us define the scope of the present study. Eight individuals were approached by email and invited to participate in a 'study into animal experiments that should rightfully be ended' (see Annex A). The format involved a semi-structured online interview of maximum one hour via Zoom or Teams, with the following key research questions: Are animal models currently in use that are actually undesirable? Why are these animal models still used? What opportunities are there to replace these animal models? Seven of the eight individuals we approached responded to our invitation, all positively. Two of the seven suggested an additional participant from their own organisation, so we interviewed these participants in pairs. In addition, we asked a number of animal rights organisations to join a group interview following this discussion. These organisations had earlier been invited to participate in a scheduled roundtable discussion with NCad. All these organisations responded positively and each sent one representative to the group interview. This group interview lasted just under an hour and 40 minutes.

The individual interviews and the group interview were recorded (with the permission of the participants) and then transcribed verbatim. The participants in one of the paired interviews objected to the recording, so notes were taken during that interview instead. The participants in the other paired interview did not object to the recording, but did object to the verbatim transcript, with as argument that they only wished to clarify outlines, not details. The verbatim transcript of this interview was therefore summarised.

Two key conclusions based on these interviews were that it will be complicated to generate a list of concrete 'undesirable' animal models or animal experiments, and that the project research question needed to be made more specific. The individual interviews and the group interview with the animal rights organisations together produced few concrete examples of animal experiments (or types of animal experiments) that deserve a place on the priority list. It was therefore decided to focus on developing well-reasoned criteria for prioritising the replacement of animal experiments (in any case for this initial phase).

A first attempt was made to conduct a systematic literature review of the academic literature, based on the keywords "priorit* AND replace* AND "animal experiment*" OR "animal research" using the search engines PubMed and Google Scholar. However, this yielded few results. A different approach was therefore adopted. Based on information obtained during interviews with stakeholders to establish the scope of the literature review, and on the knowledge and expertise of the author of this report and NCad itself (including NCad's Support Office), a list of *possible* criteria was drawn up. These criteria are as follows:

5. Degree of **translatability** of the research
6. Degree of **relevance** of the **goal** of an animal experiment
7. Degree of **distress** caused to laboratory animals
8. Degree of violation of **integrity** of laboratory animals
9. Degree of **proportionality**
10. The **stage of development** of **alternatives**
11. **Numbers of animals** or **procedures**
12. **Animal species** used

Upon further consideration, it was decided that criteria 1 to 4 were *suitable*, but criteria 5 to 8 were *unsuitable* for further exploration. The section on 'Unsuitable criteria for prioritising the replacement of animal experiments' explains why criteria 5 to 8 were deemed unsuitable for further elaboration, after which the section on 'Suitable criteria for prioritising the replacement of animal experiments' elaborates on criteria 1 to 4. Each of these latter four criteria was further explored through a narrative review using the search engines PubMed, Google Scholar and Google Search, supplemented by 'snowball sampling' and documents contributed by experts in the author's network. Within the time available, it was decided to place the most emphasis on the criteria of translatability and violation of integrity, respectively due to the increasing call to pay more attention to translatability in the ethical assessment of animal experiments (Eggel and Würbel, 2021; Garner et al., 2017; Meijboom, Kostrzewa and Leenaars, 2020; Olsson and Sandøe, 2021; Pound and Nicol, 2018; Pound and Ritskes-Hoitinga, 2018; Wehling, 2009; Würbel, 2017; Landi, Everitt and Berridge, 2021), and due to the need for further clarification of what exactly entails a violation of integrity.

Unsuitable criteria for prioritising the replacement of animal experiments

This section explains why the degree of proportionality, the stage of development of alternatives, numbers of animals or procedures, and the species of animal used constitute unsuitable criteria for prioritising the replacement of animal experiments.

Degree of proportionality

The degree of proportionality was considered unsuitable as a *separate* and *independent* criterion when prioritising the replacement of animal experiments. Proportionality is a measure of the suitability, necessity and adequacy of an animal experiment in the context of inflicted harm (Würbel, 2017), and itself refers to *other* criteria. The suitability and adequacy of animal experiments are partially described respectively under the criteria of translatability (see 'Translatability of experiments: external validity') and the relevance of the goal of an animal experiment (see 'Relevance of the goal of an animal experiment'). Necessity is additionally related to the stage of development of alternatives.

The stage of development of alternatives

The stage of development of alternatives to a particular type of animal experiment was also considered as a criterion for the replacement of animal experiments. The idea of this criterion is that priority should be given to replacing those types of animal experiments for which the development of an alternative is already in sight (leaving aside the 'scores' for other criteria), i.e. to 'harvest the low-hanging fruit'. For example, it is generally agreed that it is scientifically and technically easier to develop alternatives for toxicity studies – which are highly standardised and repetitious – than to find alternatives for hypothesis-driven academic research. Unlike the other criteria, the stage of development of alternatives is primarily a strategic criterion. This criterion in itself does not suggest that we should try to replace an animal experiment *because it poses the greatest ethical challenges*, but rather promises quicker results because an alternative is in an advanced stage of development. It was therefore chosen not to develop this criterion further. Of course, solutions in an advanced stage of development still deserve attention.

Numbers of animals or procedures

The number of animals or procedures involved is not *in itself* a good criterion for prioritising the replacement of animal experiments. Numbers of animals are relevant, but always in relation to other factors, such as distress and the research goal. For example, in 2017, the largest numbers of procedures that caused severe distress (more than 264,000) were related to batch potency testing (European Commission, 2020). This factor, however, is addressed by the distress criterion (see 'Distress in laboratory animals')

Animal species used

The primate species chimpanzee, bonobo, orangutan and gorilla have a special status in the Experiments on Animals Act. Animal experiments on these species are banned outright (Experiments on Animals Act, 2021, §10e, Article 1). Experiments with other 'non-human primates' may only be conducted under special conditions that are more stringent than for other animal species (Experiments on Animals Act, 2021, §10e, Article 2 and 3). The UK also has stricter requirements for the use of cats, dogs and equine species (Animals (Scientific Procedures) Act 1986, 2013; Schedule 2B 'Additional conditions for the grant of certain project licences'). Furthermore, in practice, research institutions choose to refrain or desist from experiments on certain animal species, as in the heart failure research at Maastricht University mentioned in the Introduction. The well-known UK Ipsos MORI survey further reveals that people think animal experiments on, for example, dogs, cats and primates are less acceptable than experiments on rats and mice (Cameron, Clemence and Xypolia, 2018).[22]

But using animal species as an independent prioritisation criterion raises other issues. The species of animal may be relevant if it involves a difference in the capacity of a species to experience distress, or differences in homological validity and thus translatability to humans, or the moral right to protection of certain species. However, further refining the criterion of a species as the *right to protection* of that species does not solve these issues. Although some people apply the social hierarchy of animal species to ethical judgements (Hursthouse, 2011; Kagan, 2016; 2018), assigning a greater right to protection based on species has also been criticised as 'speciesist' and therefore unsuitable (Olsson and Sandøe, 2021; Singer, 1993). It therefore seems there is insufficient consensus to use the right to protection of an animal species as a criterion for prioritising the replacement of animal experiments.

[22] However, Danish studies have found that animal species as a criterion carries less weight for people than, for example, the distress caused (T. B. Lund, Lassen and Sandøe, 2012; T. B. Lund et al., 2014).

Suitable criteria for prioritising the replacement of animal experiments

The literature review elaborated the following criteria for prioritising the replacement of animal experiments: the degree of translatability of the research, the relevance of the goal of an animal experiment, the distress caused to laboratory animals, and the violation of the integrity of laboratory animals. The following four sections elaborate on these criteria.

Translatability of the research

> *We believe that more rigorous and extensive evaluations of animal models are necessary. (...) there are fields where it is widely recognized that existing animal models are insufficient. (...) We propose that, for improved ethical animal experimentation, animal research in these situations should be avoided until improved models are developed. (Vieira de Castro and Olsson, 2015)*

The translatability of animal experiments has long been an issue in biomedical research (Ferreira et al., 2020; Garner, 2014; Garner et al., 2017; Greek and Kramer, 2019; A. Knight, 2019; Leenaars et al., 2019; Pound and Ritskes-Hoitinga, 2018; Ram, 2019; Van der Worp et al., 2010; Zeiss, 2015). The point here is that data obtained from animal experiments have limited translatability to humans.

The percentage of drugs that show promise in phase I clinical trials, but do not make it to market, has been between 86% and 93% for some time now (Alliance for Human Relevant Science, 2020; Garner, 2014; Hay et al., 2014; Kola and Landis, 2004; Thomas et al., 2016; Wong, Siah and Lo, 2019). Drugs typically do not make it to market because of a lack of efficacy and for safety reasons, but commercial reasons such as a lack of funding, portfolio prioritisation and intellectual property issues also play a role (Cook et al., 2014; Hay et al., 2014; Thomas et al., 2016). However, there is much more ambiguity surrounding the step from experiments with laboratory animals to clinical trials with humans. A recent review of reviews, into the translatability of findings of animal experiments into those of phase I/II trials with humans, found that the success rates range from 0% to 100% (Leenaars et al., 2019). This is not very encouraging, as it means that the studies are potentially not predictive for humans. That is, the criterion of predictive validity may not be met.

Animal models and model animals

There are other forms of validating animal models in addition to predictive validity. Before discussing the various validity criteria of animal models, however, it is worthwhile clarifying what we mean by 'animal models'. First, we distinguish between animal models and model animals (cf. Insel, 2007). A statement heard in practice such as "The animal models will be killed as part of the experiment" is an example of reification: the fallacy where an abstract construct is wielded as if it were something concrete.[23] But an animal used as a model for a human disease or condition is not itself a simplified version of that disease or condition (Hauskeller, 2007).

Since an *animal model* is an abstract model or construct, it is not something that can be killed, as a concrete experimental animal can. So, in addition to animal models, we need to distinguish concrete experimental or model animals.

It should also be noted that, when the use of an animal model is discussed in practice, it actually almost always concerns using laboratory animals as models for physiological, biological or behavioural phenomena in humans. So in this sense, there is no question of using animal models in, for example, agricultural research, because there the laboratory animal almost always belongs to the same species as the target animal. For example, new feeds for pigs (the target animals) are tested on pigs (the laboratory animals) and not on any other species.

On the other hand, it can be argued that the test pigs in the aforementioned example are also models for domesticated pigs of a particular breed or breeding line. This does not concern the extrapolation of one species to another, as in biomedical research, but rather of individual pigs in an experimental setting to the relevant breed or breeding line that is used in the target setting (here: pig farming).

In this report, we apply the commonly used concept of an animal model as an extrapolation from animals to humans. Consequently, we consider animal models to be simulations of physiological, biological or behavioural phenomena in humans using a species of animal, based on a naturally occurring, genetically modified or experimentally induced trait within that species, and with the aim of generating knowledge that can be extrapolated to the

[23] See also, for example, Held's much-quoted definition (1980), quoted in Hau and Schapiro (2011, 2): "a living organism in which a normative biology or behaviour can be studied, or in which a spontaneous or induced pathological process can be investigated, and in which the phenomenon in one or more respects resembles the same phenomenon in humans."

human target population.[24] Examples of a naturally occurring, genetically modified or experimentally induced trait are, respectively, the canine leukocyte adhesion deficiency (CLAD) model (Bauer et al., 2004), the ClockΔ19 mouse model (Kristensen, Nierenberg and Østergaard, 2018), and the experimental autoimmune encephalomyelitis model in mice, rats and macaques (Procaccini et al., 2015).

In contrast, model animals are concrete, living animals used in an experiment *to model* certain physiological, biological or behavioural phenomena in humans. The *way* they model a phenomenon is specified in the animal model. Thus, Irish Setters with CLAD in an experimental context are dogs that could be used as models for studying leukocyte adhesion deficiency in children.[25] The way they model the disease relates to the actual or hypothetical similarities between the symptoms, disease progression and mechanisms of leukocyte adhesion deficiency in dogs and humans.

Thus, the translatability of animal experiments directly concerns those research areas where animals are used for the benefit of acquiring knowledge of physiological, biological or behavioural phenomena in humans, and not, for example, veterinary research, agricultural research or fundamental research in other animal species.[26]

Classical validity criteria of animal models

"Translatability is the validity of the model for duplicating a specific aspect of targets, mechanisms, and/or disease in humans" (Landi, Everitt and Berridge, 2021). So, translatability has everything to do with the validity of animal models. An animal model (of a disease) has validity only insofar it is a reliable model of the human disease or disorder (Belzung and Lemoine, 2011). The question, of course, is when is an animal model a reliable model. Several aspects can be distinguished here, one of which is the previously mentioned predictive validity: the ability to predict a human phenomenon based on the performance of an animal model (Geyer and Markou, 1995). This may involve predicting the effect of therapies and treatments, specific markers of a disease, or disease progression (Belzung and Lemoine, 2011; Holmes, 2003; Van der Staay, Arndt and Nordquist, 2009). Predictive validity is one of the three classical forms of validating animal models (Willner, 1984) and evaluating the translatability of animal experiments (Landi, Everitt and Berridge, 2021). In addition to predictive validity, there is also face validity (phenotypic validity) and construct validity.

Face validity refers to the degree of similarity between the animal model and the disease or condition under study across the widest possible range of symptoms (Belzung and Lemoine, 2011; Geyer and Markou, 1995; Van der Staay, Arndt and Nordquist, 2009; Willner and Mitchell, 2002). A face validity study thus mainly attempts to reproduce one or more core clinical diagnostic criteria in an animal model. Although some researchers restrict the symptoms to the level of cognition, emotion and behaviour (Belzung and Lemoine, 2011; Van der Staay, Arndt and Nordquist, 2009), this is not always necessary. Similarities between symptoms may also relate to neurobiological and physiological abnormalities (Geyer and Markou, 1995; Holmes 2003; Willner and Mitchell, 2002). Examples include weight loss and increased activity in the hypothalamic-pituitary-adrenal axis in depression (Holmes, 2003; Willner and Mitchell, 2002).

Roughly three definitions of construct validity are found in the literature, with the first two probably deriving from Willner's classical formulation (1984). In the first definition, construct validity is described as the requirement that the experiment or model must measure what it is supposed to measure and can be interpreted unambiguously (Geyer and Markou, 1995; Sjoberg, 2017). The forced swim test, for example, was developed as a test for the symptom of despair in depression, but a newer interpretation is that it actually tests floating behaviour as a coping strategy (Commons et al., 2017; Molendijk and De Kloet, 2015; 2019; De Kloet and Molendijk, 2016; Molendijk and De Kloet, 2021).

The second definition of construct validity describes the extent to which an animal model is consistent with the theory underlying the modelled disease or condition in humans (Holmes, 2003; Willner and Mitchell, 2002). In the case of depression, for example, theorised neurobiological, etiological or psychological mechanisms can be tested using animal models.[27] The problem here is that, particularly in neuropsychiatric research, there is no single theory, but rather several, and consequently it is also unclear which theory the animal model should be tested against.

The third definition of construct validity focuses directly on the degree of similarity between the mechanisms underlying the disease or disorder in humans, and the observed abnormalities in the animal model (Sams-Dodd, 2006; Van der Staay, Arndt and Nordquist, 2009).

[24] This is close to Sjoberg's (2017) definition, corrected for reification in Sjoberg's definition, and partly following Hau and Schapiro's (2011) description of different types of animal models.

[25] Of course, they can also be models for dogs in general, as in veterinary research.

[26] The latter in particular can also provide essential insights into the human situation, although this was not an objective of the research. One example is the development of optogenetics, in which the response of specific cells in living tissue (such as neurons) to light is being studied in species such as fruit flies, clawed frogs and mice (for a review, see Deisseroth, 2015). The translatability criterion is not (yet) applicable in this situation, because the problem definition assumes a specific biological process, and not a specific social context such as 'the clinic'.

[27] However, sometimes etiology is described by a separate criterion, namely etiological validity (Belzung and Lemoine, 2011; Geyer and Markou, 1995).

Belzung and Lemoine (2011) also call this 'mechanistic validity'. This is in contrast with face validity, which focuses on the *effects* of the actual or hypothetical mechanisms.

Newer criteria for the validity of animal models

There have been further developments since the definitions of predictive validity, face validity and construct validity were formulated. The three classical validity criteria have been supplemented with other criteria, and partly reinterpreted. Incidentally, these criteria are not always 'new'. For example, convergent validity and discriminant validity were formulated as early as 1959 in the psychological context (D. T. Campbell and Fiske, 1959). However, these criteria were only later incorporated in the explicit evaluation of animal models (Geyer and Markou, 1995).

Validity criteria according to Garner

Garner (2014; 2017) combines the three classical criteria of validity into one – as he calls it – 'dimension'. He also distinguishes two other dimensions. The second dimension describes internal validity and external validity, and the third dimension describes convergent and discriminant validity. Garner does not provide definitions or descriptions of the various types of validity, but rather formulates key questions.

Internal validity involves the question of whether the methodology and results of the test or model are consistent with both the theory and the existing data produced by the model. This involves both reliability and replicability (or reproducibility) (Van der Staay, Arndt and Nordquist, 2009). A reliable animal model will produce consistent outcomes when repeated, and is reproducible in different contexts (i.e. laboratories). External validity involves the question of whether the results of the test or model are widely applicable (generalisability). This primarily concerns the question of whether the insights gained are applicable to the population or breeding line under study, but secondarily also how they apply at the species level and ultimately to the human situation.

Convergent validity involves the key question of whether the experiment or model is broadly consistent with the characteristics of what is being measured, or with the characteristics of the modelled human situation. Here, we want to know whether an animal model or experiment correlates with other animal models or experiments that model the same phenomenon. Finally, external validity involves the question of whether the results of the experiment or model rule out the possibility of alternative processes or diagnoses. This is the flip side of convergent validity, since in discriminant validity, a low correlation between various experiments or animal models is desirable because they are supposed to model and measure various phenomena.

Validity criteria according to Belzung and Lemoine

Belzung and Lemoine (2011) likewise formulated other validity criteria in addition to the three classical criteria, and also reinterpreted the three classical validity criteria. They distinguish between face validity, mechanistic validity, predictive validity, homological validity, and pathogenic validity.

Belzung and Lemoine define face validity according to the classical interpretation. However, they additionally specify face validity by ethological validity and biomarker validity. The former refers to the degree of similarity in behaviour between model animals and humans, while the latter refers to the degree of similarity between biotic markers. An example of the latter is the degree of similarity between abnormal concentrations of corticosterone in the blood of rats and cortisol in humans. It is clear that this does not concern the absolute concentration of these glucocorticoids in the blood (which differs between rats and humans), nor an identical chemical composition, but rather the similarity of the function of the marker: corticosterone in rats performs the same function as cortisol in humans (Belzung and Lemoine 2011).

Mechanistic validity as an interpretation of classical construct validity was discussed earlier.

Belzung and Lemoine further specify predictive validity as the similarity of the relationship between trigger factors and the occurrence of disease in model animals on the one hand (induction validity), and the therapy or medication and the disease in humans on the other hand (remission validity).

Homological validity concerns the chosen species and breeding line. An example given by Belzung and Lemoine is *Caenorhabditis elegans*, which is a poor choice to model the reduction of hippocampal volume in depression (*C. elegans* has no hippocampus!), but a better choice when focussing on basic serotonergic phenomena under stress conditions.

Pathogenic validity concerns the similarity between an animal model and the processes in humans that lead to disease. Here, Belzung and Lemoine distinguish between ontopathogenic validity and trigger validity. Ontopathogenic validity concerns the similarities between early environmental factors that make an organism vulnerable in later stages of life. Trigger validity subsequently describes the similarities between factors that make an organism pathological.

These five criteria with associated subcriteria are presented by Belzung and Lemoine as a generic framework for validating animal models of psychiatric disorders.

Table 1: *Overview of validity criteria*

Classical validity criteria	Validity criteria according to Garner		Validity criteria according to Belzung and Lemoine	
	Dimension	*Subtype*	*Criterion*	*Subcriterion*
Face validity	Face vs. construct vs. predictive vs. vs.	Face validity	Face validity	Ethological validity
				Biomarker validity
Construct validity		Construct validity	Mechanistic validity	-
Predictive validity		Predictive validity	Predictive validity	Induction validity
				Remission validity
	Internal vs. external	Internal validity		
		External validity		
	Convergent vs. discriminant vs.	Convergent validity		
		Discriminant validity		
			Homological validity	Species validity
				Breeding line validity
			Pathogenic validity	Ontopathogenic validity
				Trigger validity

Scope of validity criteria

These validity criteria were developed for research into neurological and psychiatric disorders such as depression, anxiety and schizophrenia (Sams-Dodd, 1999; Willner, 1984; Willner, Muscat and Papp, 1992). Assessments of animal models using validity criteria have also been largely limited to this field of research (e.g. Belzung and Lemoine, 2011; Garner, 2014; Meijboom, Kostrzewa and Leenaars, 2020; Morrice, Gregory-Evans and Shaw, 2018; Treit, Engin and McEown, 2010; Tricklebank and Garner, 2012). Although Sams-Dodd (2006) developed a generic framework for evaluating the clinical validity of disease models, their examples are also limited to neuropsychiatry. Denayer et al. (2014) looked beyond neuropsychiatry and applied a modified version of Sams-Dodd's validity criteria framework to the field of oncology.

It is unclear why the application of various validity criteria has been limited to neuropsychiatry. One possibility is that neuropsychiatry faces the greatest translatability problems. To date, insufficient information has been collected to be able to assess whether there are indeed significant differences in translatability between the various fields of research (Leenaars et al., 2019). What is known, however, is that oncology drugs have the lowest success rate in clinical trials (93-97% failure rate, see Hay et al., 2014; Thomas et al., 2016; Wong, Siah and Lo, 2019), and not drugs for neurological and psychiatric disorders.[28] This would suggest that, rather than neuropsychiatry, it is oncology that faces the greatest translatability problems (cf. Mak, Evaniew and Ghert, 2014). This may explain why the validity criteria are now also applied within oncology. On the other hand, cardiovascular research also scores low success rates (93% failure rate; Hay et al., 2014; Thomas et al., 2016)[29], but the application of validity criteria in this field has been limited to date.

In any case, there are no reasons in principle *not* to apply validity criteria to a broader range of fields than just neuropsychiatry and oncology. In fact, a criterion such as face validity can implicitly play a role, for example in infection research. One component of the search for a valid animal model for SARS-CoV-2 infections in humans involved the degree of similarity between clinical signs such as nasal discharge, coughing, increased body temperature, and elevated cytokine profiles in the blood (Bertzbach et al., 2021; A. Singh et al., 2020; Zhai et al., 2021). Singh et al. (2020) therefore suggest validating animal models for SARS-CoV-2 using, for example, the three classical validity criteria (cf. Swearengen, 2018). More recent tools

[28] Therapies such as various forms of radiotherapy do not fall under the FDA classification of 'drugs' and are therefore not included here. Immunotherapy is included, as it involves drugs such as anti-PD-1 checkpoint inhibitors (durvalumab, pembrolizumab, nivolumab) and anti-CTLA-4 antibodies (ipilimumab).

[29] It is worth noting that, based on other publications, the drug failure rate for cardiovascular disease is many times lower, namely 74-80% (Kola and Landis, 2004; Wong, Siah and Lo, 2019). This may have to do with the way success rates are calculated. See Wong et al. (2019) for a discussion of the methodology.

designed by Ferreira et al. (2019; 2020) and Wehling (Wehling, 2009; Wendler and Wehling, 2012; 2017) to estimate the validity of animal models consequently have a generic focus (see 'Tools for determining the validity of animal models').

It should also be noted that translatability is a relevant criterion wherever an animal model is involved (see also 'Animal models and model animals'). Translatability is also important for fundamental research in the biomedical domain, for example into the underlying mechanisms of MS. However, this concerns the translatability of the *knowledge* of a specific biological or etiological phenomenon, acquired through an animal model, to the human situation. In applied research, however, it concerns the translatability of the *effectiveness* of a technical product, such as an implant or a drug. Fundamental and applied research in the biomedical domain thus have different focuses – which can also be combined in an individual research project – but translatability is relevant to both.[30]

Translatability as criterion

In the literature, internal validity is often discussed in the context of the *experimental design*: is it randomised, is it blind, has a power calculation been performed beforehand, have appropriate control groups been selected, etc. (Henderson et al., 2013)? These are important procedures in keeping with the principles of good science practice to ensure the reproducibility – and reliability, in the sense that this says something about the measuring instruments used – of animal experiments. However, reproducibility is not only guaranteed by randomisation, blinding, etc. (i.e. good science practice); it also depends on the method by which a specific animal model is generated. For example, the translatability of a multiple sclerosis (MS) study in rhesus monkeys based on the experimental autoimmune encephalomyelitis (EAE) model depends not only on an accurate power calculation, randomisation and blinding, but also on the reproducibility (and reliability) of the EAE model itself.[31] The latter is determined by the method used to generate the relevant animal model. So, internal validity involves both a general experimental aspect, and an aspect specific to the animal model used.

The same applies for external validity. Ferreira et al. (2020) point out that the external validity criterion does not only concern the trial design (such as including relevant endpoints), but also the animal models themselves. Henderson et al. reach a similar conclusion:

[30] Fundamental biomedical research is therefore a good example of 'oriented basic research', as distinguished from 'pure basic research' as in the example of optogenetics in footnote 6 (Eurostat, 2021).

[31] External validity also plays a role here.

> *While threats to external validity can be addressed by replication studies under varying experimental conditions, threats to construct validity are reduced by articulating, addressing, and confirming theoretical presuppositions that underly clinical generalisation. (Henderson et al., 2013, 2).*

The external validation of animal experiments thus involves general experimental aspects, such as choosing relevant endpoints and conducting replication studies under varying experimental conditions (cf. Voelkl et al., 2021).[32] This increases the likelihood of generalisability of the animal experiment in question. However, this generalisability also depends on the animal model itself: the three classical validity criteria (face, predictive and construct validity), and newer validity principles (such as homological validity), are what allow for the assessment of the latter aspect.[33,34] This is summarised in Figure 1.

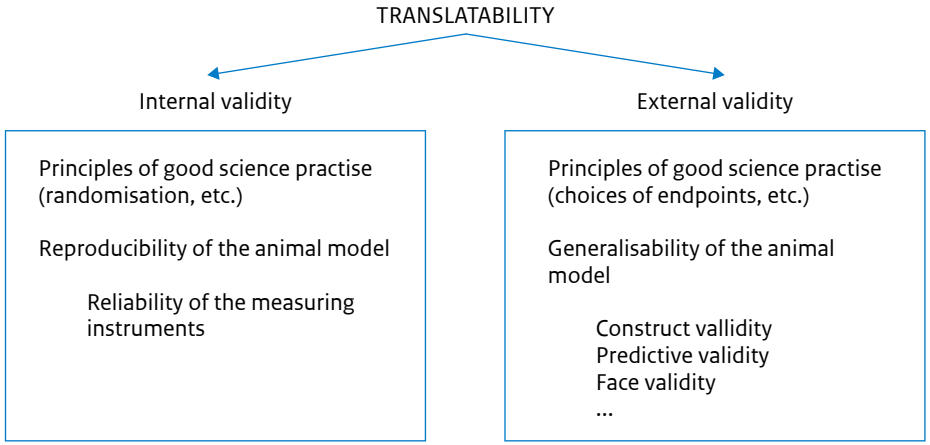


Figure 1. *Relationships between validity criteria in the translatability of animal experiments*

[32] The latter is relevant, for example, to correct for the high degree of genetic homogeneity in breeding lines of experimental animals such as mice and rats, as opposed to the wide genetic variety within the human population.

[33] Although Garner identifies several dimensions of validity, he does not clarify how these dimensions relate to each other. This is important where translatability is used as a criterion for prioritising animal experiments. After all, we want to know if the various validity criteria all carry equal weight, and if an animal model should ideally score well for each validity criterion. Or if there is some form of hierarchy in the various validity criteria. Based on the findings of Garner's report, the latter seems possible.

[34] According to Pound and Ritskes-Hoitinga (2018), the general experimental aspects concern 'surmountable problems' where it involves improving the translation of the results of animal experiments, and it is the more theoretical aspects of the animal model, and particularly differences between species, that generate 'insurmountable problems' (cf. PETA 2020).

The aim of this report is to establish useful criteria for prioritising the replacement of animal experiments. One of these criteria is the translatability of an experiment. Translatability is dependent on the internal and external validity of animal experiments. Internal and external validity both involve a general experimental aspect and an aspect specific to the relevant animal model. This implies two things for prioritising the replacement of animal experiments in practice:

1. Animal experiments that do not meet the general experimental requirements of research, such as randomisation and blinding, are generally considered undesirable. General experimental requirements are important for ensuring good science practice, and guidelines that prescribe such requirements should therefore be endorsed by every scientific institution (cf. Henderson et al., 2013; Hooijmans, Leenaars and Ritskes-Hoitinga, 2010). Yet they are not applied as criteria when setting priorities for the replacement of animal experiments, precisely because they are so generic to all experiments. In other words, general experimental criteria cannot differentiate *between* animal models.
2. This is different for the animal model itself. The translatability of an animal model depends on reproducibility (internal) and generalisability (external). Here, the reproducibility of an animal model partly depends on the method used to build a specific animal model, while the generalisability of an animal model partly depends on the extent to which the model simulates the presumed underlying mechanisms of a biological or etiological phenomenon in humans.[35] Reproducibility and generalisability thus differ at the generic level per animal model as subcriteria of translatability, and so can be used as subcriteria of translatability when prioritising the replacement of animal experiments (see also Figure 1).

Tools for determining the validity of animal models

An important question is how to determine the validity of animal models in general, and separately by subfield (phenotype, mechanism, etc.). Willner (1991) proposed that this was a value judgment and not something that can be measured. Recently, however, tools have been developed to determine the validity and translatability of animal models in a structured manner. These are the models of Van der Staay, Arndt and Nordquist (2009), Sams-Dodd (2006) (further elaborated by Denayer et al. (2014)), Ferreira et al. (2019; 2020) and Wehling (Wehling, 2009; Wendler and Wehling, 2012; 2017). Van der Staay, Arndt and Nordquist's tool is a *qualitative* tool, in the sense that the final assessment of an animal model using predefined criteria is consensus-based. In the tools of Sams-Dodd/Denayer et al., Ferreira et al., and Wehling, the final assessment is based on the *quantitative* score assigned to an animal model using predefined criteria. The tools are briefly explained below.

[35] This makes construct (or mechanistic) validity the most scientifically important criterion, but at the same time it is the most difficult aspect of an animal model to assess (Holmes, 2003). The latter is the reason why face validity is often used, as symptoms can be compared between model animals and human patients.

Tool developed by Van der Staay, Arndt and Nordquist (2009)

Van der Staay, Arndt and Nordquist (2009) have developed a flowchart in which the scientific evaluation begins with determining the reliability and reproducibility of an animal model (internal validity), followed by the face validity, predictive validity, construct validity and external validity (see Figure 2).

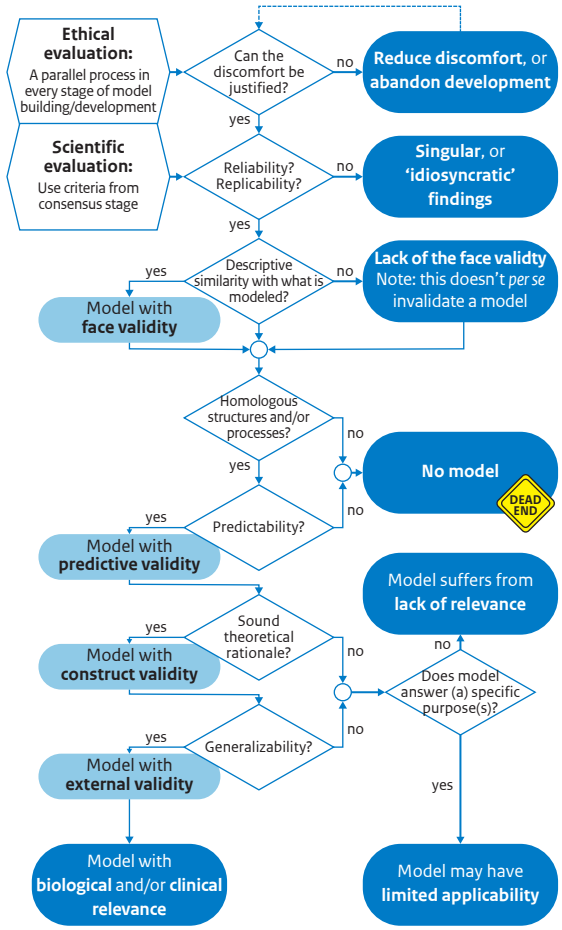


Figure 2. *Flowchart for evaluating an animal model. Reproduced from Van der Staay, Arndt and Nordquist (2009) based on Creative Commons Attribution License CC by 2.0.*

Tool developed by Sams-Dodd (2006)/Denayer et al. (2014)

Partly due to the fact that the classical use of the various validity criteria does not allow for the graduated assessment of animal models (see, for example, Van der Staay, Arndt and Nordquist's flowchart discussed above) and it is not possible to compare different models (e.g. in vivo versus in vitro versus in silico), Sams-Dodd (2006) proposed a graduated assessment of different models.

Sams-Dodd works with three variables: the organism used (e.g. mouse or human), the complexity of the structure (e.g. cell or organ or whole organism) and the way the disease develops (e.g. induced by chemical administration or 'naturally' by genetic predisposition). Sams-Dodd goes on to score validity in terms of none, low, medium, and high. For example, the use of primates scores 'medium' for organism, cell lines score 'low' for complexity, and experiments with human patients scores 'high' for disease induction.

To arrive at an overall assessment of the three variables, Sams-Dodd developed an algorithmic scoring system. For example, the use of tissue from mice *ex vivo*, where the disease is induced by chemical means, is given an overall score of 3 because it deviates from the clinical situation for all three criteria (organism is mouse rather than human, complexity is tissue rather than the whole organism, and disease induction is artificial rather than natural/actual). The highest score (0) is given for experiments on a human patient (human, intact organism, actual disease progression). All other possible combinations score between 0 and 3.

Denayer et al. (2014) further developed Sams-Dodd's algorithmic scoring system by adding the criteria of face validity and predictability (i.e., predictive validity), and increasing the highest possible score per criterion to 4. For face validity, for example, no similarity in symptoms between clinical practice and the animal model gives a score of 1, one similarity a 2, one core symptom a 3, and more than one core symptom a 4. The higher the score across all criteria, the greater the likely validity of the animal model.

Tool developed by Ferreira et al. (2019; 2020)

Because classical validity criteria are multi-interpretable, Ferreira et al. (2019; 2020) say they have developed an objective scoring system to compare the validity of different animal models. They believe the existing system developed by Sams-Dodd and Denayer et al. overlooks several aspects, such as histology and biomarkers. Based on a literature review and interviews with stakeholders, Ferreira et al. propose a 'Framework to Identify Models of Disease' (FIMD) encompassing eight domains: epidemiology, symptomatology and natural history, genetics, biochemistry, etiology, histology, pharmacology, and endpoints (see Table 2). Closed core questions (yes/no) are formulated for each of these domains. A predefined score is awarded to each positive answer, and the maximum score per domain is the same for all domains. The higher the score for all criteria, the greater the likely validity of the animal model.

Table 2. The Framework to Identify Models of Disease (FIMD). Reproduced from Ferreira et al. (2019) based on Creative Commons Attribution License CC By 4.0.

	Weight
1. Epidemiological validation	12.5
1.1 Is the model able to simulate the disease in the relevant sexes?	6.25
1.2 Is the model able to simulate the disease in the relevant age groups (e.g. juvenile, adult or ageing)?	6.25
2. Symptomatology and natural history validation	12.5
2.1 Is the model able to replicate the symptoms and co-morbidities commonly present in this disease? If so, which?	2.5
2.2 Is the natural history of the disease similar to human's regarding:	
2.2.1 Time to onset	2.5
2.2.2 Disease progression	2.5
2.2.3 Duration of the symptoms	2.5
2.2.4 Severity	2.5
3. Genetic validation	12.5
3.1 Does this species also have orthologous genes and/or proteins involved in the human disease?	4.17
3.2 If so, are the relevant genetic mutations or alterations also present in the orthologous genes/proteins?	4.17
3.3 If so, is the expression of such orthologous genes and/or proteins similar to the human condition?	4.16

	Weight
4. Biochemical validation	12.5
4.1 If there are known pharmacodynamic (PD) biomarkers related to the pathophysiology of the disease, are they also present in the model?	3.125
4.2 Do these PD biomarkers behave similarly to humans?	3.125
4.3 If there are known prognostic biomarkers related to the pathophysiology of the disease, are they also present in the model?	3.125
4.4 Do these prognostic biomarkers behave similarly to humans?	3.125
5. Etiological validation	12.5
5.1 Is the etiology of the disease similar to humans?	12.5
6. Histological validation	12.5
6.1 Do the histopathological structures in relevant tissues resemble the ones found in humans?	12.5
7. Pharmacological validation	8. 12.5
7.1 Are effective drugs in humans also effective in this model?	4.17
7.2 Are ineffective drugs in humans also ineffective in this model?	4.17
7.3 Have drugs with different mechanisms of action and acting on different pathways been tested in this model? If so, which?	4.16
8. Endpoint validation	12.5
8.1 Are the endpoints used in preclinical studies the same or translatable to the clinical endpoints?	6.25
8.2 Are the methods used to assess preclinical endpoints comparable to the ones used to assess related clinical endpoints?	6.25

Tool developed by Wehling (Wehling, 2009; Wendler and Wehling, 2012; 2017)

The last scoring system, Wehling's model, focuses on the early assessment of the translatability of pharmacological research (Wehling, 2009; Wendler and Wehling, 2012; 2017). Five dimensions are distinguished: starting evidence (e.g. what is already known from *in vitro* data?), human evidence (e.g. from clinical trials); biomarkers for predicting efficacy and safety (e.g. how 'close' to the clinical picture is a biomarker?); *proof-of-mechanism*, *proof-of-principle* and *proof-of-concept* testing (e.g. clarity and acceptance of surrogate endpoints); and aspects linked to personalised medicine (e.g. what is known about the impact of genetic variation on the pharmacokinetic and pharmacodynamic processes). The subcomponents of the various dimensions all have different weights, and in the most recent publication this is specified by disease area. For example, based on the latest findings, Wendler and Wehling (2017) assign the greatest weight to the 'biomarkers for predicting efficacy and safety' dimension in oncology research, but in psychiatric research, the 'human evidence' dimension has the greatest weight. The weight per subcomponent is multiplied by a score between 1 and 5, where 1 stands for weak and 5 for strong. Wehling (2009) thereby provides applicable characteristics for weak and strong scores respectively. After adding up the individual scores, this number is divided by 100, giving a final score between 1 and 5. A score of 4 or higher represents good translatability.

Choice of tool

For the purposes of this report, it would go too far to compare the various tools for determining the validity of animal models and suggest the use of a particular tool. The aim was to provide an overview of tools that could be helpful for prioritising the replacement of animal experiments with regard to translatability. Ideally, the selection of tools for assessing the translatability of animal experiments in practice will be refined in consultation with scientists specialised in this field.

Is translation of animal experiments even possible?

Since the publication of LaFollette and Shanks' 'Brute Science: Dilemmas of Animal Experimentation' (1996), even the *very possibility* of translating insights gained through an animal model to humans has been fundamentally called into question. LaFollette and Shanks point out that it is insufficient to only present successful or failed cases of translation – with proponents of animal experiments presenting successful cases and opponents presenting failed cases – to be able judge if translation is possible. Indeed, negative results are inherent to conducting research. It is necessary to delve deeper into what failed cases of translation actually mean: do they reflect a methodological defect, or are they part of 'normal scientific research' (Kuhn)? LaFollette and Shanks believe that such a methodological defect is indeed present in laboratory animal research. There is an ontological problem that exists in causal disanalogies between model animals and humans: causal mechanisms are almost never identical between model animals and humans. According to LaFollette and Shanks, animal models can therefore only serve as *hypothetical animal models* and not as *causal animal models*. LaFollette and Shanks thus argue that causal disanalogies preclude the translational step from model animals to humans, and draw the conclusion that animal models can at most serve to generate hypotheses for human research.[36] This also calls into question the ethical acceptability of animal experiments.

There is also an epistemological issue: we cannot determine whether an animal model has external validity if we do not already know exactly what we hope to learn from the extrapola-

[36] The professional literature also touches on this point of ontological differences, for example in A. Knight (2019) and Pound & Ritskes-Hoitinga (2018).

tion. In other words, according to LaFollette and Shanks, we can only find out whether the ontological claim of comparability is met by comparing the results of animal research with the results of human research.[37] This is known as the 'extrapolator's circle' (Steel 2010).

While it goes too far to discuss this matter in detail here, it should be noted that LaFollette and Shanks' argument has not gone unchallenged. Nordgren (2010), for example, cites Steel's work on *comparative process tracing*, which says that a causal mechanism need not be completely identical for a model animal and a human, but that we need to look at the similarities between the phases of this causal mechanism (cf. Steel, 2010). Nordgren also suggests that a combination of *in vivo*, *in vitro* and *in silico* methods could reinforce the potential for extrapolation, and that using only non-animal methods also has serious limitations. Finally, Nordgren notes that LaFollette and Shanks' approach is too theory-driven and does not do enough justice to the pragmatic and provisional nature of scientific practice. More specifically, Degeling and Johnson (2013) criticise LaFollette and Shanks' distinction between causal animal models and hypothetical animal models, saying it devalues hypothetical animal models and betrays LaFollette and Shanks' ethical agenda. Degeling and Johnson instead propose making a distinction – that already exists in everyday laboratory animal practice (Hau and Schapiro, 2011) – between exploratory, explanatory and predictive animal models. This latter distinction also recognises the validity criteria of face, construct and predictive validity, respectively. So, based on Degeling en Johnson's proposal, the criterion of translatability can also be used to prioritise the replacement of animal experiments.

Summary

Translatability concerns the extent to which the data obtained from an animal experiment are translatable to humans. Where the starting point of the problem definition of a research proposal is a social practice such as medicine, translatability is a relevant criterion for both the applied and the fundamental research within that domain. The applied research then concerns the translatability of a product: is a drug or device that is effective in a laboratory animal also effective in humans? The fundamental research involves the translatability of knowledge of a specific mechanism to the clinical situation: does the model have clinical relevance?

Translatability also concerns validity: an animal model (of a disease) has validity only insofar as it is a reliable model of the human disease or disorder, and more specifically insofar as the same underlying mechanisms are involved. Experimental aspects such as randomisation and blinding are typically cited in the context of validating animal experiments. While these quality requirements are obviously important – animal experiments that do not meet high scientific standards are always undesirable – these requirements do not say anything about the usefulness of the animal model *itself*, even if the experiment is conducted in accordance with high scientific standards. However, reproducibility and generalisability as subcriteria of validity *are* relevant for assessing the animal model itself. *Reproducibility* indicates whether the model produces consistent and reliable results, i.e. are the right things being measured? *Generalisability* indicates whether the results obtained with the animal model can be extrapolated to the human situation, i.e. does the animal model indeed constitute a good model for the phenomenon we want to study in humans? The three classical criteria of face, construct and predictive validity are generalisability criteria.

Several tools have been developed to estimate or measure the validity of animal models. Preferably, a single tool, or several suitable tools, will be selected in consultation with the professional field.

Animal experiments that use animal models with a low degree of translatability have first priority for replacement, also depending on the assessments of other criteria, such as the relevance of the goal of an animal experiment.

[37] This is particularly problematic for LaFollette and Shanks, because they say that scientists consider human research to be "scientifically second-rate" (LaFollette and Shanks, 1996, 23).

Example of translatability

FORCED SWIM TEST

The forced swim test (FST) has been used in research into depression since the 1970s. The test involves placing a rodent in a container of lukewarm water from which it cannot escape. The animal will actively swim around the container and try to escape, but will give up at some point and passively float. This duration of immobility is the primary outcome parameter of the test. Because clinically used antidepressants reduce this duration of immobility when applied in rodents, a shorter duration of immobility is interpreted as evidence of the efficacy of newly tested drugs (Porsolt, Le Pichon and Jalfre, 1977).

The test is the subject of heated debate, mainly because of its limited translatability to humans in terms of depression (for a review, see Trunnell and Carvalho, 2021). Whether the FST can be assigned face validity depends on the interpretation of the rodents' floating behaviour. If this is interpreted, according to Porsolt, Le Pichon and Jalfre (1977), as an expression of despair (a sign of depression in humans), then the FST has face validity. However, this interpretation is contested; in the alternative interpretation that FST tests floating behaviour as a coping strategy, it does not have face validity for measuring depression (Commons et al., 2017; Molendijk and De Kloet, 2015; 2019; De Kloet and Molendijk, 2016; Molendijk and De Kloet, 2021). Furthermore, the FST has no etiological validity, because the 15-minute (rat) or 6-minute (mouse) test differs significantly from the slow pathogenesis of depression in humans (Molendijk and De Kloet, 2015). Moreover, antidepressants are given to healthy animals, which does not reflect clinical practice. Closely related to this is the lack of construct validity, because the underlying mechanisms of depression are unknown (Sewell et al., 2021). It is strongly debated whether the FST has any degree of predictive validity (high degree: Slattery and Cryan, 2012; none: Trunnell and Carvalho, 2021). So, based on the criterion of translatability, the FST used in depression research is an example of an experiment that has high priority for replacement. In fact, this is already happening in research institutions and the industry (PETA, 2022; Ritskes-Hoitinga, personal communication, 30 September, 2021).

Relevance of the goal of an animal experiment

Animal experiments are conducted with a goal. A distinction is made between the immediate and the ultimate goal (CCD, 2019).[38] The immediate goal is the goal that can be answered within the context of an animal experiment, while the ultimate goal lies further ahead in the future. For example, the immediate goal of Alzheimer's research could be to further unravel the mechanism underlying the formation of amyloid-beta plaques between nerve cells in the brain. The ultimate goal is then to contribute to preventing, curing, or alleviating Alzheimer's disease. This concerns the multitude of experiments and research projects carried out within a scientific field that together can substantially contribute to this ultimate goal.

Types and variants of animal experiments

The various ultimate goals that animal research can serve involve two factors. The first is the *type* of goal pursued by the research. A precedent was established in 2013 with a ban on animal experiments for cosmetic goals (although exceptions still exist) (European Parliament and the Council, 2019). The interests involved in animal experiments for cosmetic goals also score low in surveys and studies, and not only in a European context (Cameron, Clemence and Xypolia, 2018; S. Knight and Barnett, 2008; T. B. Lund et al., 2014; Sandgren et al., 2020). These surveys also reveal that respondents rate safety testing of chemicals for use in household and commercial contexts only slightly higher than cosmetic testing. Research for the benefit of human and animal health consistently scores highest (see also Ormandy and Schuppli, 2014).

Sandgren et al.'s (2020) survey of faculty staff and students at the University of Wisconsin-Madison in the United States also included agricultural research as a category.[39] Respondents considered such animal research more justifiable than cosmetic research, but less justifiable than research for the benefit of human and animal health. However, the category of agricultural research itself does not explain everything. It can, after all, be assumed that the degree of acceptance of agricultural research varies depending on the *variant* of agricultural research. For example, in 2018, at the behest of the Central Authority for Scientific Procedures on Animals (CCD), the then State Secretary for Economic Affairs of the Netherlands, Van Dam, commissioned an advisory report from the Council on Animal Affairs (RDA) on the desirability of using animal experiments for research into conventional and intensive livestock farming. The RDA (2018b) advised that animal experiments in livestock farming should ideally contribute to the health and welfare of livestock and the transition to a sustainable livestock industry.

[38] These goals are closely related to the benefits of the research (Niemi, 2021). In analogy with the distinction between immediate and ultimate goals, Brønstad et al. (2016) speak of actual and promised benefits. It would go beyond the scope of this report to elaborate on the exact relationship between benefits and goals here.

[39] It should be kept in mind that such surveys among faculty staff and students have serious limitations in terms of generalisability to a wider audience (Ormandy and Schuppli, 2014).

This implies that, for example, research into the use of regionally sourced protein sources in the diets of dairy cows is more justifiable than research into the inclusion of a tropical crop in the feed of sows in farrowing pens. More in general, the degree of acceptance of animal research therefore also depends on the variant used within a given type of research. This is the second factor.

Assessing different types and variants of animal experiments

Variants of animal research are also rated differently within medical research. For example, surveys have revealed that people consider the use of laboratory animals for the purposes of research into diseases like Alzheimer's and cancer more acceptable than research into obesity (T. B. Lund, Lassen and Sandøe, 2012; T. B. Lund et al., 2014). One aspect that plays a role in respondents' answers is that obesity is a condition that is, in principle, preventable (through lifestyle adjustments) and that people perceive preventing obesity as an individual responsibility (Freriks et al., 2005; T. B. Lund, Lassen and Sandøe, 2012). Other factors include the mortality and morbidity associated with a disease, and the extent to which existing means are available to make living with a disease or condition bearable. For example, migraine is accorded less weight in this regard than Alzheimer's disease and cancer (T. B. Lund, Lassen and Sandøe, 2012).

However, the example of obesity as a so-called lifestyle disease also demonstrates why this is a complicated discussion. As Freriks et al. (2005, 68) rightly observe with regard to the question of whether animal experiments should be allowed for developing drugs for managing overweight and obesity:

> *At first glance, it seems it is easy to provide the morally appropriate answer. Animal welfare organisations will answer this question in the negative, but other parties will tend to nuance their answer. Not all cases of obesity, they will argue, can be blamed on individual responsibility, so it would be unfair to deny the 'innocent' victims of this condition a potential cure just because others with the same ailment have only themselves to blame.*

On top of this, social and addiction factors can also play a role. This conundrum also emerged in the debate early this millennium on whether animal research into conditions linked to smoking is ethically justifiable. In its 2003 annual report, the then Food and Consumer Product Safety Authority reported on a case that played out within an animal ethics committee (VWA 2004) involving research into the harmful effects of aldehydes in cigarette smoke. Such research is (potentially) "ethically problematic, because smoking is a bad habit that is avoidable and socially contested" (Freriks et al., 2005, 69). However, this too can be qualified, because "in the case of an addiction-disease such as smoking, which is called a 'lifestyle' disease, to what extent can one still speak of an individual responsibility?" (Schurgers, 2005, 35). Moreover,

> *To what extent can people blame a disease on their own deliberately risky behaviour anyway It is important to realise here that, during sensitive periods in life such as puberty, the reasons for starting smoking can be different to when someone takes up the habit in adulthood. And, of course, there is also the question of passive smokers. If the animal ethics committee 'decides' that animal experiments cannot be carried out for the benefit of smokers, it also deprives the latter two groups of the prospect of a treatment (Schurgers, 2005, 37).*

While animal rights organisations may be quick to rule that such research into lifestyle-related conditions is unacceptable, a nuance will always be found to explain why such an animal research variant should nevertheless go ahead (Freriks et al., 2005).

Another difficulty regarding the definition of the relative value of research has to do with the fact that insights obtained from one type of research, or a variant of a type of research, may have relevance for a broader field. Thus, animal experiments into the etiology of obesity may also provide insights that are relevant to a broader group than only people with obesity. So perhaps animal research for the benefit of human nutrition should be ranked even higher than medical research, because nutritional interventions can be used to strengthen the immune system and thus potentially help protect against all kinds of diseases and disorders (so-called nutraceuticals).

Another example is research into breeding eels in captivity (CCD, 2021). The current eel farming sector depends on glass eels caught in the wild. These glass eels are fattened for sale to the consumer on eel farms. The big problem with eels is that they are both an endangered species and do not naturally reproduce in captivity. However, a hormone treatment has been developed to obtain eggs and roe from parent animals and develop them into larvae. If this method of raising glass eels is successful, it will serve both the goal of conserving the species and improving the sustainability of eel farming. The problem here is that the conservation of the species can be attributed great value, while the production of glass eels for fattening in eel farms may be valued less (the consumption of eels is unnecessary and there are also alternatives). But both these outcomes fall under the goals of the research.

More in general, Olsson et al. (2015) conclude that the relative importance of different (ultimate) goals of research is difficult to evaluate, because so many issues are involved and

there is no agreement on how the various goals should be ranked. Also, the European working document on 'Project Evaluation and Retrospective Assessment' further argues that there is a lack of agreement on the relative importance of using animals for, for example, educational purposes, statutory safety testing, human health or animal health, and therefore no simple hierarchical order can be given (Expert Working Group for Project Evaluation and Retrospective Assessment, 2011).

It is true that it is difficult to define a simple and precise hierarchy of the ultimate goals of animal research. Nevertheless, a trend can be discerned. VanDeVeer's (1995) distinction between basic, serious and peripheral interests can be helpful here. *Basic interests* involve specific conditions or goods that are required by humans and animals to enable "minimally adequate" functioning, such as food, water, oxygen or the absence of chronic, intense pain (VanDeVeer, 1995, 88).[40] For example, hunting seals represents a basic interest for the traditionally living Inuit, while for people in modern agricultural regions, keeping animals for slaughter does not. *Serious interests* concern goods or conditions without which a human or animal can survive, but only with difficultly or at a cost. For example, chickens have a serious interest in dust bathing and pigs in rooting. *Peripheral interests*, finally, concern a good or condition without which life is possible, but the human or animal cannot flourish without it. Examples include the cultural use of decorations such as ivory pendants and, to some extent, the use of grooming products.

If we translate the weights of these interests into goals of animal research, research with 'basic goals' could involve studies aimed at preventing, curing and making more tolerable life-threatening diseases and diseases that seriously impair the quality of life of humans[41] and animals, and the conservation of animal species and ecosystems. Serious goals could include studies aimed at developing sustainable forms of agriculture and animal husbandry, and safety testing of chemicals needed for sustainable food production. Examples of peripheral goals might be studies into the safety of cosmetic products (already banned in the EU), fur production (banned in the Netherlands since 2021), production goals of intensive livestock farming, and safety tests for chemicals for use in the plastics industry.

The relative importance of these goals of research must be seen in the light of today's modern society, and against the background of the use of laboratory animals. All the goals of research are legitimate in themselves, but a certain hierarchy can be defined in *relation to the use of laboratory animals*. For example, it is possible to conceive a situation where, in an age and society where there are no alternatives to wearing fur, and people are dying of cold, research into fur production becomes a basic goal.

To be able to weigh the goals of animal research and the associated values, it is also important to take our current knowledge as starting point, and only make plausible assumptions. So, it is reasonable to argue that research into the etiology of obesity will be of primary benefit to those suffering from this condition. Any possible wider applications of insights gained from such research fall under the category of 'flukes' (serendipity), and the small chance of that happening should not be given any weight when defining the importance of the goal of a proposed animal experiment.

Of course, research can serve several purposes at once, as the eel example made clear. However, this in itself does not form an obstacle for defining a hierarchy of the goals of research. A practical solution here is to always start with the most 'serious' goal when assessing whether animal experiments should be replaced sooner for a particular form of research. So, forms of research with a more 'serious' goal, will have lower priority for replacement (also depending, of course, on how they score on other prioritisation criteria).

Finally, it should be kept in mind that this concerns defining *prioritisation criteria* in the search for replacements for animal experiments, and not, for example, the desirability of an outright ban. This means that the lack of precision when establishing the importance of the categories of research goals forms only a relatively minor problem.

Summary

Animal experiments are conducted with a goal. The immediate goal of a concrete animal experiment (e.g. unravelling a mechanism) can be distinguished from the ultimate goal (e.g. a therapy for multiple sclerosis) which lies further in the future and requires multiple experiments and studies. Some of those ultimate goals are the subject of debate. For example, animal research for the purpose of manufacturing cosmetics has been banned in Europe and many people also question whether it is still necessary to carry out chemical safety tests on animals.

The ultimate goals of research can be ordered based on VandeVeer's distinction into basic, serious and peripheral interests. Research with a basic goal could involve studies aimed at preventing, curing and making more tolerable life-threatening diseases and diseases that seriously impair the quality of life of humans and animals, and the conservation of animal species and ecosystems. Serious goals could include studies aimed at developing sustainable

[40] Such conditions are specific to a species or breed: for example, dairy goats require a housing in the winter, while many sheep breeds do not.

[41] We do not make a further distinction between lifestyle-related diseases and other diseases or conditions here, because acquiring a disease or condition is always partly determined by the environment. Moreover, there will always be more and less culpable groups of people who will benefit from the outcomes of animal research.

forms of livestock farming and food safety tests. And in today's modern society, livestock research to improve production could be called a peripheral goal.

To be able to weigh the goals of animal research and the associated values, it is also important to take our current

These can all be legitimate goals of research, but a hierarchy of importance can be established when the use of laboratory animals is required to achieve such a goal. Projects and animal models that primarily serve peripheral goals should have highest priority for replacement.

Example of the relevance of the goal of an animal experiment

POST-WEANING DIARRHOEA IN PIGLETS

Research has been conducted into ways to address health problems in piglets for more than two decades now. This focuses on piglets that suffer diarrhoea during the period shortly after being separated from their mothers. A range of factors make piglets sensitive to infections which often result in diarrhoea during the period after weaning. Particularly now that the downsides of using antibiotics on livestock have become clear, much of the research has been focused on preventing post-weaning diarrhoea, such as using improved feed and adjusting management and housing systems (see for reviews: J. M. Campbell, Crenshaw and Polo, 2013; Rhouma et al., 2017; Wensley et al., 2021).

Although the studies have already led to various adjustments in farming practice, post-weaning diarrhoea remains a major problem for pig farmers. This gives rise to the following questions: To what extent have the animal experiments carried out thus far contributed to solving the problem? Are proposals for research into the effects of yet another feed supplement not merely 'sticking plasters'? And are there not other, more fundamental alternatives available (for example, leaving piglets with their mothers for longer or keeping them in lower densities)? Another option would be to opt for more sustainable solutions, although these would require the entire sector to change in the long term.

So how should such animal research be judged in the light of the relevance criterion? According to the distinction between basic, serious and peripheral interests, this type of research into post-weaning diarrhoea in piglets can be said to have a peripheral interest: it concerns research that primarily serves an economic interest, and only to a very limited extent piglet health. Moreover, alternatives are available. So, replacing such animal research therefore has high priority.

This example is reproduced with permission in modified form based on: RDA (2018b).

Distress in laboratory animals[42]

Animal research involves distress, in other words damage to the welfare of laboratory animals. Given the definition of an animal experiment in the Experiments on Animals Act, all animal experiments inherently involve at least mild distress.[43] Within the European Union, the distress suffered by laboratory animals is the main form of damage covered by the mandatory harm-benefit analysis. In addition, Europe prescribes an upper limit to harm:[44]

> *Subject to the use of the safeguard clause in Article 55(3), Member States shall ensure that a procedure is not performed if it involves severe pain, suffering or fear that is likely to be long-lasting and cannot be ameliorated (European Parliament and Council of the European Union, 2010, Article 15.2).*[45]

The said safeguard clause states that long-lasting severe distress may be permitted by individual member states in exceptional and scientifically justifiable cases.

European directive 2010/63/EU also prescribes a classification of the severity of procedures in terms of non-recovery, mild, moderate, and severe distress. The severity of procedures is "determined by the degree of pain, suffering, distress or lasting harm expected to be experienced by an individual animal during the course of the procedure" (European Parliament and Council of the European Union, 2010, Annex VIII). Although the European directive describes pain, suffering and distress in the same breath, Olsson et al. (2019) argue that a qualitative distinction should be made between different degrees of pain (and illness) on the one hand, and suffering (and distress) on the other. The degree of pain depends on the duration and intensity, but the crucial question is whether the animal is capable of *coping*. The moment the pain becomes overwhelming for the animal, "compensation cannot occur, normal life cannot be experienced, and) the animal cannot fully recover and will be fundamentally changed even if the external situation improves" (Olsson et al., 2019, sec. 'How to Measure Severe Suffering). This may be the reason why the European directive has also included a separate clause on long-lasting severe distress, in addition to the mandatory harm-benefit analysis.

In the context of prioritising the replacement of animal experiments, theoretically, animal experiments that involve severe and prolonged distress (suffering) will be high on the agenda. However, animal experiments involving severe and prolonged distress are not carried out in the European Union. If a Member State wishes to carry out an animal experiment with severe and prolonged distress, it must report this to the European Commission within one month. However, Olsson et al. (2019) report that no notification of an animal experiment involving severe and prolonged distress had been received by the European Commission as of 1 July 2019. While this may be reason for cautious optimism, it should be kept in mind that the estimation of the degree of distress involved is the responsibility of individual committees, and there are only general rules for this (Olsson et al., 2019). So there may be some variation in how the rules are interpreted and applied. Consequently, the absence of serious and prolonged distress may be merely an administrative description. In any case, an enquiry should be conducted to establish if the European Commission received any notifications in recent years.

After severe and prolonged distress, animal experiments that involve any period of severe distress should logically have next priority for replacement. In the UK, 4% of animal experiments conducted in 2020 involved severe distress (Home Office, 2021). This was 0.9% in the Netherlands in the same year (NVWA, 2022). According to the latest European figures however, in 2016 and 2017, 11% of animal experiments involved severe distress (European Commission, 2020).

Of further significance here is that, from a European perspective, so-called *batch potency testing* is by far the largest category of severe distress, followed at some distance by nervous system experiments and diagnostic tests (European Commission, 2020).

So, the scale used in the European Union to 'measure' distress can be used to prioritise the replacement of animal experiments. This prioritisation ranges between severe and prolonged, severe, moderate, and lastly mild distress.

Summary

Animal research involves distress, in other words damage to the welfare of laboratory animals. European directive 2010/63/EU prescribes a classification of the severity of procedures in terms of non-recovery, mild, moderate, and severe distress. This classification of distress can be used to prioritise the replacement of animal experiments, starting with the promotion of alternatives to animal experiments and animal models that involve (long-lasting) severe distress.

[42] This section has been kept short because of a parallel NCad study into severe distress and the already wide knowledge of measuring distress in the field.

[43] See the Introduction for the definition.

[44] Recently, DeGrazia and Beauchamp (2020) also launched a framework of six criteria to determine the moral acceptability of animal experiments, one of which concerns an upper limit to harm.

[45] See also Experiments on Animals Act (2021, §3, Article 10b(2)).

Example of distress

ROUTINE BATCH TESTING FOR CLOSTRIDIUM VACCINES

In Europe, batch potency testing constitutes by far the largest category of animal testing with severe distress. The European Pharmacopoeia requires that every batch of a vaccine produced must be tested for efficacy. Little is known about which specific vaccines involve severe distress in laboratory animals. A 2008 study by the RSPCA provides the most, albeit dated, information (J. Cooper and Jennings, 2008).

One of the examples that gives rise to the greatest concerns about distress in laboratory animals involves the efficacy testing of batches of veterinary vaccines (J. Cooper and Jennings, 2008; Cruelty Free Europe, 2022; Hendriksen, personal communication, 9 February 2022). An example of this is the Clostridium vaccines. Concerning vaccines against blackleg (Clostridium chauvoei) in sheep and cattle, and botulism (Clostridium botulinum) in sheep, cattle, horses and birds, the Pharmacopoeia is quite unrestrictive and allows exposure tests to be conducted (J. Cooper and Jennings, 2008). It is these exposure tests in particular (in addition to toxin-neutralising tests) that cause a lot of distress during routine batch testing. Mice (C. botulinum) and hamsters (C. chauvoei) are divided into vaccinated and unvaccinated groups and then exposed to the respective pathogen. The laboratory animals suffer severe distress during these tests. In the case of C. chauvoei, this involves oedema formation, tissue necrosis and ultimately death of the hamsters in the unvaccinated group (K. P. Singh, Parihar and Tripathi, 1992, quoted in J. Cooper and Jennings, 2008). In the case of C. botulinum, this involves dehydration, loss of sensorimotor reflexes and cardiac or respiratory failure in mice in the unvaccinated group (Luvisetto et al., 2003, cited in Cooper and Jennings, 2008).

Thus, in regard to distress, replacing routine batch testing of Clostridium vaccines has high priority. Importantly in this respect, replacement in vitro tests are under development for both C. botulinum (Rust et al., 2017) and C. chauvoei (Nicholson et al., 2019).

Violation of the integrity of laboratory animals

Background to the concept of integrity

Several authors have argued that the animal ethics debate would benefit from an explicit discussion about the violation of the integrity of laboratory animals in research projects (De Vries, 2004; Röcklinsberg, Gamborg and Gjerris, 2014; Verhoog, 2004). The Central Authority for Scientific Procedures on Animals has also included the integrity of laboratory animals as an evaluation item in the advice form that animal ethics committees submit to the Authority (question C12) (see also CCD, 2019).[46] Although it was already applied in laboratory animal practice, the concept of 'animal integrity' first emerged in the Netherlands in the 1990s as part of the debate about interventions such as dehorning cattle, docking tails of pedigree dogs, and the genetic modification of animals (Brom, 1997; De Vries, 2009). In this regard, the concept of integrity violation refers to morally problematic issues that lie 'above' or 'behind' welfare violations (Bovenkerk, Brom and Van Den Bergh, 2002; Röcklinsberg, Gamborg and Gjerris, 2014; De Vries, 2006). Even though an operation like docking a dog's tail is virtually painless when done under anaesthesia and according to good veterinary practice, many people perceive such operations as morally problematic. The concepts of integrity and integrity violation provide an entry point for the discussion about this issue.

Most studies that discuss integrity refer to Rutgers and Heeger's 1999 definition (Bovenkerk, Brom and Van Den Bergh, 2002; Gjerris and Gamborg, 2010; Grimm, 2014; Nordgren, 2010; Ortiz, 2004; Schmidt, 2008; Verhoog, 2007).[47] This classical definition of animal integrity is as follows: "*[the] wholeness and completeness of the animal and the species-specific balance of the creature, as well as the animal's capacity to sustain itself independently in an environment suitable to the species.*" (Rutgers and Heeger, 1999, 45).[48] Three concepts play a central role in this definition: (1) the wholeness of the animal; (2) the species-specific balance of the animal; and (3) the ability of the animal to sustain itself independently in an environment appropriate for the species. Rutgers and Heeger cite the dehorning of cattle as an example of the violation of the wholeness of the animal. Regarding the violation of an animal's species-specific balance, Rutgers and Heeger give the example of breeding for rapid muscle growth in broilers, which eventually restricts the animal's capacity to walk and sometimes even results in spontaneous death. This is because the metabolism is out of balance in these broilers. Finally, an example of a violation of the animal's capacity to sustain itself independently in an environment appropriate for the species is the breeding of 'double-muscled' cattle like the Belgian Blue (Rutgers and Heeger, 1999, 47-48). These animals show exceptionally strong muscle growth but have a narrow birth canal, a physical configuration that has been achieved through domestication and selection and does not occur in cattle bred in a more natural context. As a result, the calves of double-

[46] In Denmark, the concept of integrity was used in 2006 in legislation governing animal cloning (Gjerris and Sandøe, 2006). It is unknown to us if it is still used.

[47] Although the concept of 'animal integrity' has strong roots in the Dutch discussion, it was used as early as 1995 in a document by a Swiss ethics committee, without any reference to Dutch authors (*Ethik-Studienkommission des Eidgenössischen Volkswirtschaftsdepartements zur Gentechnologie im ausserhumanen Bereich*, 1995). This suggests that 'animal integrity' as a concept does not have purely Dutch origins.

[48] The Council for Animal Affairs (RDA) more or less adheres to the following definition of integrity: "Integrity encompasses the species-specific wholeness and completeness of an animal and its ability to function independently, according to its nature, in an environment appropriate to its species" (RDA, 2018a, 25).

muscled breeds cannot be born *naturally* or in a *species-specific* way, but must always be delivered by a caesarean section. So, these double-muscled cattle are always dependent on intensive assistance from humans.

The examples provided by Rutgers and Heeger themselves to illustrate integrity violations all relate to farm animals and companion animals. There are also examples of integrity violations in *laboratory animals* according to Rutgers and Heeger's definition (this does necessarily entail agreement with Rutgers and Heeger's definition at this point). A violation of the wholeness of the animal, firstly, could be toe or ear clipping in genetically modified mice for the purposes of identification and genetic characterisation (cf. NCad, 2020). A highly controversial 'example', secondly, of violating the species-specific balance of laboratory animals, is the mouse with a human ear on its back (cf. Trouw, 1995). A more subtle example is the prevention of horn formation in cattle using CRISPR technology. Finally, an example of a violation of an animal's capacity to sustain itself independently in an environment appropriate for the species are the nude mice that are regularly used in oncological research. These mice lack a thymus and fur, and so require a highly controlled laboratory environment (both microbiologically and climatically) to be able to function biologically and mentally. This makes them strongly dependent on humans.

In summary, in the now classical definition of Rutgers and Heeger, integrity concerns the wholeness of the animal, the species-specific balance of the animal, and the ability of the animal to sustain itself independently in an environment appropriate for the species.

Reception of Rutgers and Heeger's definition of integrity

With regard to the reception of Rutgers and Heeger's definition of integrity, there are three points that should be noted. The first concerns the interpretation of the first subcriterion, the wholeness of the animal. The second concerns a shift in the interpretation of integrity towards a stronger focus on the capacities or abilities of animals. The third is a critical note regarding the third subcriterion of functioning independently in an environment appropriate for the species. These three points are discussed separately below.

Wholeness of the animal

The question is what exactly should be understood by the 'wholeness of the animal'. Is dehorning cattle problematic because it directly intervenes in the animal body, or is it problematic because the cattle then lack a body part that they naturally should have?[49] Brom describes wholeness as the "wholeness of the animal in the physical and physiological sense" (Brom et al., 1996, 15). However, his interpretation is very broad, as evidenced by his example of a nude mouse that lacks the fur that it should naturally have (see also 'Example of integrity: nude mice').[50] This is in line with the view that dehorning cattle is problematic because the animals lack something that they should naturally and physically have as a species.

Schmidt (2008, 319), however, argues that operationalising the wholeness of the animal firstly and foremostly concerns *crossing the boundaries of the animal body*, as in the case of an amputation.[51] This is in line with the view that tail docking is problematic because it involves a direct intervention in the animal body. Breeding a nude mouse, for example, would not constitute a violation of animal integrity, because no literal physical boundaries are crossed (see also 'Integrity of the individual, breeding line or breed, and species').

So, there are at least two possible interpretations of the wholeness of the animal. This is discussed further in the section on 'Narrow and broad interpretations of integrity'.

Animal abilities

As early as 1996, Brom wrote that, in addition to the wholeness of the animal in the physical and physiological sense (see the section on 'Wholeness of the animal' above), integrity can also be seen as (1) "the ability to function species-specifically and to achieve this ability" (Brom et al., 1996, 15) or (2) "the ability of the animal to maintain a physiological and ethological balance in a specific environment without intensive human assistance" (Brom et al., 1996, 15). The first interpretation "mainly concerns ethological and functional factors. Here it refers to an 'ethogram' describing all the more or less essential species-specific behavioural expressions of

[49] Or: is it problematic because the cow lacks certain abilities as a result of dehorning? This is addressed under 'Animal abilities'.

[50] 'Naturally' refers to ideas such as 'naturalness' and '*telos* ' (Hauskeller, 2005; Kramer and Meijboom, 2021), and involves all manner of interesting relationships between the concepts of integrity, naturalness and *telos*. For example, Hauskeller says that "*biological integrity consists in the ability to live according to one's natural ends [or]* telos" (Hauskeller, 2016, 47). This needs to be explored further, however not in the context of the present report.

[51] A legal question here is what the term 'crossing the boundaries of an animals body' contributes to the existing regulatory framework, that defines an animal experiment as involving "experimental or other [and] educational goals, which may cause the animal a level of pain, suffering, fear or lasting harm equivalent to, or higher than, that caused by the introduction of a needle in accordance with good veterinary practice" (Experiments on Animals Act, 2021, §1, Artikel 1a). Does the current definition of what constitutes an animal experiment by law not already cover any violation of the wholeness of the animal? It is in any case so that the wholeness of the animal, as understood in a *narrow* sense, is by definition violated during an animal experiment. However, this says nothing about the *degree of violation* of this wholeness/integrity (see 'Integrity: absolute or relative?'). Furthermore, the legal definition of an animal experiment has a different relationship with the concept of 'wholeness' in a *broader* sense.

an animal" (Brom et al., 1996, 15).[52] Animals should have the opportunity to exhibit species-specific behaviour. Pigs should be able to root in the mud, chickens to scratch, and rodents to dig tunnels. This aspect of the opportunity to exhibit species-specific behaviour has also been given more weight in the recent amendment to the Animals Act (House of Representatives, 2021). The second interpretation does not refer to the species, but rather to the degree to which the animal can independently and successfully 'cope', and so is close to certain views regarding animal welfare (e.g. Broom, 1993; Korte, Olivier and Koolhaas, 2007).

In line with this focus on *abilities* of animals, Heeger and Brom somewhat later came with another definition of integrity: "*the physical and mental ability of the animal to realise their well-being and to flourish*" (Heeger and Brom, 2001, 245). In general, abilities can be said to be "that which a human (or non-human animal) is or could be capable of under the right conditions" (Nussbaum 2006, 119).[53] So, here, animal abilities concern that which enables an animal to achieve well-being and flourish. Here, Heeger and Brom come very close to the definitions of well-being in terms of coping as discussed above. This subsequently raises the question of the extent to which integrity actually still says anything about problematic issues 'above' or 'behind' animal welfare.[54]

De Vries (2009, 88) points out that Heeger and Brom's definition of integrity does not completely coincide with the classical definition: for example, the dehorning of cattle is problematic according to the classical view of integrity because the wholeness of the animal is violated. However, according to Heeger and Brom's definition, such an intervention is only problematic if it violates the ability of the animals to achieve well-being and flourish.[55] According to the newer definition, violation of the (physical) wholeness of the animal is not always a form of violation of the animal's integrity. So, the newer definition does not define 'wholeness' in absolute terms. Instead, it takes into account that which the animal has the *potential* to achieve, based on its genome or a specific body part, to determine whether an intervention is an integrity violation.[56]

Independent functioning

As mentioned earlier, the third subcriterion of Rutgers and Heeger's definition involves the ability of the animal to sustain itself independently in an environment appropriate for the species. Schmidt elaborates this subcriterion theoretically – and originally – by arguing that integrity does "not only mean a *state* in which [the animals] can find themselves, but also involves a specific *ability to integrate*" (Schmidt, 2008, 318, translated from German to Dutch by CJR). The animal itself *actively* maintains the interaction between the various parts of its body,[57] and the incorporation of its body as a whole into the environment (*Umwelt*) (Schmidt, 2008). A concrete example is nest-building in birds:

> *Building a nest consists of a pattern of different behaviours, which in turn form part of a larger pattern of behaviour including social behaviour between male and female birds, breeding behaviour, and the pair caring for their youngsters. These behaviours form part of the behavioural pattern of the relevant bird species. Thus, in addition to the species-specific wholeness of organs, the wholeness of a particular bird consists of its ability to organise its own behaviour (Lijmbach, 1996, 68, emphasis added).*

[52] Cf. "When essential species traits have disappeared in an animal, the integrity of that animal is said to have been violated" (RDA, 2016, 37).

[53] Although the discussion of integrity in terms of abilities contains no explicit reference to Nussbaum's abilities approach (as far as we know), it does show formal similarities. Nussbaum translates ten key human abilities into their animal context, namely the ability to live, healthiness, bodily integrity, enjoyment, experiencing emotions through engagement with others, practical thinking (to the extent this ability is present), establishing social bonds, relationships with other species, play, and having control over the environment (Nussbaum, 2006; Janssens, 2020). It is interesting that Nussbaum includes 'bodily integrity' as an ability. She seems to mean this in the sense of 'the ability to maintain bodily integrity'. Yet bodily integrity would appear to be a more fundamental trait, in the sense that it is a prerequisite for having any abilities at all, as has been discussed in the animal integrity debate (cf. Schmidt, 2008).

[54] As mentioned above, Heeger and Brom explicitly mention the relationship with animal welfare when they state that integrity comprises "*the physical and mental ability of the animal to realise their well-being and to flourish.*" Since Heeger and Brom consider 'respect for integrity' to be a separate requirement to a 'good life' for animals, this implies that they do not reduce integrity violations to welfare violations.

[55] De Vries claims that dehorning cattle does not adversely affect their ability to live a good life and flourish, and therefore dehorning does not violate the integrity of the animal. But this is debatable. Welfare violations in the form of pain at the location where the horns are burned away, for example, can be long-lasting, which is a violation of the animal's ability to have a good life.

[56] De Vries ultimately opts for the newer definition because (among other reasons) he thinks it can do justice to almost all moral intuitions and judgements that might be involved in animal integrity issues.

[57] Verhoog (2007, 369) says that "*integrity presupposes the existence of an 'organism', a living whole with interconnected parts. It is the interconnectedness, the balanced harmony of the parts of the whole, which is somehow linked to the concept of integrity*". Others also note that integrity implies 'wholeness', but thereby also an integration of 'parts' within a given 'whole' (Holdrege, 2002; quoted in Hauskeller, 2016; Schmidt, 2008; Nordgren, 2010). We speak of 'wholeness' when these various parts exist in a balanced relationship with each other. Seen in this way, the first two subcriteria of the classical definition are two sides of the same coin: the 'wholeness' in the first subcriterion automatically assumes the 'balance' in the second subcriterion. The death of an animal is thus the ultimate violation of its integrity, because "*the whole is completely broken into pieces and ceases to exist*" (Nordgren, 2010, 70).

While at an initial level of integrity violation, this would involve crossing the animal's more or less static physical boundaries (see above, 'Wholeness of the animal'), at the next level it indicates the active maintenance by the animal of the boundaries of its body in relation to its environment. It is this 'active' component that is referred to in the third subcriterion in Rutgers and Heeger's definition.

The question is whether, by stating that animals should be able to maintain themselves *independently,* Rutgers and Heeger's definition does not actually go further than this active component. This definition would appear to be too absolute. Firstly, this makes the concept of integrity difficult to explain in relation to domesticated animals, that by definition no longer live in their 'original' natural habitat (Brom, 1997, 139). Brom (1997, 139) therefore suggests a more manageable criterion: "the extent to which an animal requires humans to satisfy its needs and, for example, be fed and kept by artificial means."

Secondly, according to the ethics of care in relation to the *human* situation, dependency and vulnerability – whether or not they are mutual – are characteristic of human existence and not necessarily problematic (e.g. Comstock, 1992; Kittay, 2011; Engster, 2019). Another relevant question is why the animals must be able to function independently; dependency relationships can have a legitimate place too, depending on the situation (Engster, 2006). It can be argued, however, that there is a moral difference between actively taking certain abilities away from humans and animals (and thereby making them more dependent), and the simple existence and acknowledgment of (mutual) dependency relationships. In the former case, there is a risk of instrumentalisation and the violation of the *dignity* of humans and animals (Balzer, Rippe and Schaber, 2000; Ortiz, 2004; Nussbaum, 2006; Hauskeller, 2007). The second case concerns a given reality.

Narrow and broad interpretations of wholeness

As mentioned earlier, there are at least two possible interpretations of the wholeness of an animal in relation to this subcriterion. In the first interpretation, wholeness (and thereby also integrity) is understood in a *narrow sense* to involve crossing the animal's physical boundaries. In this sense, for example, toe clipping – where it has no further consequences for the mouse's locomotion and other behaviour – involves a violation of animal integrity, but not the breeding of nude mouse lines that are severely limited in their ability to adapt to the microbial environment and climate.[58] The latter case does not involve a direct physical intervention.

Wholeness, however, can also be interpreted within a *broader* interpretation of integrity, involving animal abilities on the one hand and the species to which it belongs on the other. In short, then, an integrity violation is *any intervention that results in a loss of species-specific abilities*. Integrity thus involves *possessing the set of abilities characteristic of the species.*[59] This explicitly concerns *the set*, in other words the combination of abilities characteristic of a species. Some of these abilities will be widely shared, but others may be exclusive to a species. For example, a more or less species-specific ability of rodents is gnawing relatively hard (organic) objects, while the ability to reproduce is a widely shared ability within the animal and plant kingdom. Lammerts van Bueren et al. (2003) refer to the 'integrity of life' for this latter case.

Because these concern abilities, the animal is required to act to *achieve* those abilities.[60] Abilities therefore 'require' an appropriate context within which they can be achieved. Laying hens in cages, for example, lack an appropriate context in which to achieve dust bathing as a species-specific behaviour. Similarly, individual housing of male mice deprives them of the ability to groom each other as a species-specific activity.[61] However, it should be noted that the failure to provide an appropriate environment for animals in these examples does not in itself constitute a violation of their integrity. The animals retain their abilities, only they are not offered an opportunity to achieve these abilities. While this is a concern,[62] it is not a violation of integrity.

Important here is that this broad interpretation of integrity can do justice to the moral unease evoked by procedures such as breeding nude mouse lines and the use of vasectomised male mice. These practices are not necessarily problematic according to a narrow interpretation of

[58] A more complicated example involves cutting off a piece of a fish's fin (fin clipping) (Berger Eriksen et al., 2011). This involves violating the wholeness of the body, but not permanently, as a fish fin grows back.

[59] Thus, the distinction made between indicators of species-specific behaviour and indicators of integrity in the RDA's advice (2016) on breeding and reproductive techniques is not very useful. Indeed, the RDA fails to give any concrete indicators of integrity. The example that it gives of chickens bred for blindness could just as easily be approached in terms of the opportunity to exercise species-specific behaviour (scratching, etc.).

[60] Nussbaum (2006) calls these 'functionings'.

[61] In the same line, stereotypies can be seen as unnatural behaviour that, in a sense, mirrors a natural or species-specific behaviour: for example, bar biting in sows is seen as surrogate rooting and nest-building behaviour. Stereotypies are repeated sequences of behaviour with no apparent function (Broom, 1991). Research into preventing stereotypies and encouraging species-specific behaviour by enriching the cages of laboratory animals such as mice has been limited to date but is currently ongoing (Bailoo et al., 2018; Gross et al., 2012; Mason et al., 2007; Novak et al., 2016).

[62] This point is addressed within broader interpretations of animal welfare like 'natural living' (Fraser et al., 1997b; Fraser, 2008; V. Lund, 2006; Green and Mellor, 2011).

integrity.[63] However, operationalising a broad interpretation of integrity will require much knowledge about the species-specific abilities of animals and the extent to which these may be violated (more on operationalisation under 'Operationalising integrity').

On the other hand, the narrow interpretation of integrity is the easiest to operationalise, and is in line with how the bodily integrity of human beings is enshrined in the Dutch Constitution and applied in human medical ethics.

Empirical utility of the concept of integrity

In the scientific and ethical literature, there is and has been much debate about whether integrity is an empirically useful concept. The question is whether integrity describes an *ideal* or that it (also) relates to *empiricism*. To discuss this question, we will start by explaining the different points of view in this debate and then move on to a more critical discussion of them, and finally propose a constructive interpretation of integrity.

One reason to deny that integrity has a relationship with empiricism was prompted by the claim that a veterinary intervention in an animal – such as a caesarean section of an ewe because the lamb is poorly positioned – does not involve a violation of the integrity of the animal, as opposed to an intervention for utilitarian purposes – such as a caesarean section of an ewe to simulate the effects of premature birth (Vorstenbosch, 1993; Rutgers and Heeger, 1999). In the first case, the intervention is primarily in the interest of the health of the ewe and lamb themselves. In the second case, the intervention is primarily in the interest of humankind: the ewe and lamb serve as a model for premature birth in humans. So, according to Vorstenbosch and Rutgers and Heeger, the *intention* of the intervention plays a role in our judgement of whether that intervention violates the integrity of an animal.[64]

A problem with this interpretation is that the result of two actions with different intentions can physically produce the same result. Docking a dog's tail for aesthetic reasons produces the same physical result – a dog without a tail – as docking a tail for veterinary reasons (Bovenkerk, Brom and Van Den Bergh, 2002). According to the interpretation of Vorstenbosch and Rutgers and Heeger, docking for aesthetic reasons would violate the integrity of the dog, but docking for veterinary reasons would not. But in that case, should we not draw the conclusion that:

> *Integrity is not a biological aspect of the animal itself after all. The concept then loses its objective, biological character and becomes a moral rather than an empirical notion. It does not refer to a notion of factual intactness or wholeness so much as to a perceived intactness. It refers to how we feel an animal should be (Bovenkerk, Brom and Van Den Bergh, 2002; emphasis in original)?*[65]

In line with this, the CBD notes that:

> *The extent to which the biological traits of an animal are violated or altered can be considered an indicator of the degree of violation of the animal's integrity. In philosophical terms, the concept of integrity violation would then be operationalised by considering the extent to which certain biological traits – such as appearance, behaviour and self-reliance – are changed. However, the two cannot be equated: integrity is evaluative, while an animal's biological traits can be objectively described (CBD, 2008, 22).*

Although this is not made explicit, integrity in this sense is considered to be *intersubjective*. In contrast to objectively observable traits such as appearance, behaviour, etc., integrity concerns a valuation of these objectively observable biotic phenomena that has wide support and is established in a dialogue between humans.

But there are several criticisms of this line of thought. To start with, this view of integrity links the intentions of the intervention to the judgement of whether a particular animal intervention violates the integrity of an animal. De Vries (2009, 89-90) argues, in our opinion correctly, that the violation of the integrity of an animal does *not* need to be linked to the intentions of

[63] At the same time, it must be noted that such procedures may also be considered problematic on other grounds, for example based on the notion that human beings belong to a given order of reality that they can only manipulate to their detriment, and must therefore practise modesty and avoid conceit (Brom, 1997, 152-56; D. E. Cooper, 1998; Scruton, 1998).

[64] Another, more peripheral, reason for arguing that integrity is not an empirical concept has its origin in Vorstenbosch's (1993) observation that an animal's integrity can only be harmed by *human* intervention. As explained by De Vries, Vorstenbosch probably does not mean to deny that the (bodily) wholeness of an animal can be altered by biotic causes, such as the occurrence of cancer or by a predator, for example, but that we only speak of a violation of animal integrity if the wholeness of the animal is violated by humans. However, it is not immediately obvious, for example, why a mouse's integrity is not compromised if its paw is torn off by a cat.

[65] Incidentally, Rutgers and Heeger themselves believe that integrity *can* be determined objectively, because it refers to "biological traits that are characteristic of the species in question" (Rutgers and Heeger, 1999, 49). One can only guess how Rutgers and Heeger intended to maintain this objective character of integrity combined with the intention of the intervention.

the intervention.[66] A veterinary intervention can also be seen as a violation of the integrity of the animal. The wholeness (physical or otherwise) of the animal is violated and/or the animal loses certain abilities (it loses a leg, is made infertile, etc.). However, such an intervention may be justified because the principle of beneficence dictates that the health and welfare of the animal comes first. In this case, the act of beneficence outweighs the necessity to respect the integrity of the animal. So, there is a moral difference between these two interventions (a veterinary intervention and a utilitarian intervention) that is not determined by the intention of these interventions.

It can also be noted here that linking intentions to the concept of integrity does not improve its empirical usefulness. The example of the caesarean section for veterinary reasons makes this clear. Caesarean sections are in the interest of the health of ewes and lambs, but also in the interest of the farmer (who wants living, healthy animals). A veterinary intervention may thus be embedded in a utilitarian practice. If we include intentions in our judgement of whether – and to what extent – the integrity of an animal is violated, in the case of a caesarean section on a sheep farm, it will not be clear whether the integrity of the ewe is violated or not.[67]

A second criticism regarding this line of thought relates to the CBD's comment that integrity has an (intersubjective) evaluative character, while biological concepts are objective (and descriptive).[68,69] However, nobody would debate that health (and illness) is a biological concept. Yet not only is it a descriptive concept – 'this rhesus monkey is healthy' – but it is also evaluative. Health is therefore sometimes referred to as a *'thick concept'* (Haverkamp, Bovenkerk and Verweij, 2018). The state of 'health' involves a certain norm; when this is deviated from, we speak of disease. Opinions differ as to what this norm means, as is illustrated by the many definitions of health, but this does not change the fact of the norm itself.

Important in this context is that animal welfare is *also* a 'thick concept': *"Any conception of animal welfare inherently involves values because it pertains to what is better or worse for animals"* (Fraser et al., 1997a, 188, emphasis in original; cf. Fraser, 1995; Tannenbaum, 1991). Animal welfare does not only concern the current state of the animal – for example, whether it is free from hunger and thirst, has comfortable shelter, is healthy, has the ability to exercise species-specific behaviour, and is free from fear[70] – but also assumes that, for example, the chronic or occasional occurrence of hunger and thirst involves distress for the animal, and is therefore problematic and undesirable.

This demonstrates that biological concepts are not purely objective-descriptive, as stated by the CBD. The proposition put forward here is that integrity is a 'thick concept' of the same sort as health and well-being (cf. Visser and Verhoog, 1999, iii). Integrity has both an empirical (or descriptive) and an evaluative side. Here, as with the concepts of health and well-being, there is room for debate about what the norm of integrity comprises (see, for example, CBG, 2002). However, this is no reason to discard the concept, any more than we need to discard the concepts of health and well-being.

Given that contributing to the welfare of laboratory animals has been widely supported and operationalised – think of the 3Rs of replacement, reduction and refinement (e.g. Fenwick, Griffin and Gauthier, 2009) – it can be argued that, by analogy, there need not be a fundamental objection to applying the concept of integrity as a prioritisation principle *simply because* it is

[66] A similar conclusion can be reached about Bovenkerk, Brom and Van den Bergh's (2002, 18-19) example of human integrity. They are reluctant to draw the conclusion that a transgender person's bodily integrity would be violated if they decided to change their gender through a medical intervention. This is because Bovenkerk et al. say that, when an intervention is desired by the affected individual themselves, there can be no violation of integrity. Again, it could be argued that such an intervention in fact does concern a violation of bodily integrity, but that this violation must be weighed against the (potential) benefit that the transgender person will find more harmony with themselves. So, we do not need to conclude that "integrity as a moral notion can therefore be diametrically opposed to integrity as an empirical notion" (Bovenkerk, Brom and Van Den Bergh, 2002, 19), because the concept of integrity has both an empirical and evaluative dimension.

[67] This is probably why Vorstenbosch concludes that "it is unusual – at least – to say that the integrity of the animal is violated when human (veterinarian) action, for instance surgery, is taken that is intended to benefit the animal – and only the animal" (Vorstenbosch, 1993, 111; emphasis added). Only if an intervention were entirely for the benefit of the animal, and not also for the benefit of, say, the farmer or researcher, would there be no violation of integrity of the animal. However, this means the concept has very little empirical usefulness, because in practice, the goals and interests of humans and animals are often intertwined.

[68] Another criticism, albeit of lesser importance, is that the CBD gives 'self-reliance' as an example of a biological trait. However, self-reliance is also very much an evaluative concept, so it is not a particularly suitable example for the CBD's argument.

[69] Pairs of concepts – such as intrinsic and extrinsic value, final and instrumental value – may be primarily subjective or inter-subjective in nature, although this is also contested (e.g. Van den Heuvel, Nullens and Roothaan, 2017). Be this as it may, integrity does need to be distinguished from a concept such as intrinsic value, but this is not always the case (see for an example: COGEM, 2018).

[70] The famous five freedoms of animal welfare as formulated in 1965 by a British commission (Rogers Brambell et al., 1965).

evaluative in nature. Of course, the use of the degree of violation of integrity as a prioritisation principle can be objected to on other grounds. One objection, for example, is that integrity is too absolute a concept and does not allow for gradation. This objection will be addressed in the 'How can integrity be operationalised?' section.

The conclusion of this section is that integrity does indeed also have an empirical component. This conclusion can be upheld if the integrity violation is not determined by the intention of the intervention, but rather has a relationship to the consequences for the animal. At the same time, integrity as a concept has an evaluative function, as do the concepts of health and animal welfare. So, there is no reason to exclude animal integrity as a prioritisation principle for the replacement of animal experiments because this concept would only be evaluative in nature.

Operationalising integrity

Integrity: absolute or relative?

Alongside the debate on the empirical usefulness of the concept of integrity, the question of whether integrity is an absolute or relative concept has also been played out in the scientific literature. Vorstenbosch suggests that integrity is a kind of yes/no concept, just like the concept of being a virgin or not (Vorstenbosch, 1993). Similarly, Bovenkerk, Brom and Van den Bergh (2002) argue that integrity appears to be an absolute concept, like being pregnant or not. In other words, there is integrity, or there is not. There are no gradations in between.

However, others have argued that there are degrees of integrity violation, and thus a relative concept of integrity. According to Rutgers and Heeger (1999), more or less severe violations of integrity involve violations of (1) the wholeness of the animal; (2) the species-specific balance of the animal; and (3) the ability of the animal to sustain itself independently in an environment appropriate for the species. It is not entirely clear whether the degree of violation of integrity is determined here by *how many criteria* are violated (with each criterion having a yes/no status), or whether the degree of violation can be determined for each of these three criteria, which would then give a cumulative degree of integrity violation.

De Vries also believes that it is possible "*to objectively assess the seriousness of the violation of integrity*" (De Vries, 2009, 91). According to De Vries, the severity of the violation is a function of the degree to which the animal's abilities to achieve welfare are reduced. For example, he argues that amputating a leg could well be a less serious violation of a dog's integrity than docking its tail, because the tail has an important social function.

However, it is important to note that both Rutgers and Heeger and De Vries refer to the degree of *violation* of integrity (see also Brom et al., 1996). However, this is not yet the same thing as degrees of *integrity*. In fact, Bovenkerk, Brom and Van den Bergh (2002) argue that while degrees of *violation* of integrity exist, degrees of *integrity*, as mentioned above, do not.[71]

However, one can ask how meaningful it is to separate the reason or cause of the integrity violation from its impact on the animal in this way. A 'violation' is always a violation of something, in this case the integrity *of the animal*. So the intervention cannot be separated from the consequences for the animal. If we want to use the degree of 'more and less' in relation to the violation of animal integrity, we will still have to consider the empirical effects on the animals.[72]

Indeed, Brom previously said that "gradation is possible because it is possible for an animal whose integrity has been violated, to have its integrity violated again. The integrity of a dog whose tail has been docked is violated, but can be violated again by ear cropping" (Brom, 1997, 131-32). In this sense, it differs from being a virgin or pregnant: once you have been deflowered, you cannot be deflowered again, and someone cannot (normally) become pregnant if they already are.

[71] Unfortunately, Bovenkerk, Brom and Van den Bergh do not elaborate on this issue, but you can argue that they have hereby disregarded the consequences of the intervention and focus instead only on the intervention itself. This is known in ethics as a deontological form of moral reasoning, as opposed to, say, a consequence-based ethical form of reasoning. This is in keeping with their linking integrity to the intention of the intervention. Yet this, too, is not entirely consistent with the example cited by Bovenkerk, Brom and Van den Bergh (2002) – amputating a dog's leg versus docking its tail – in which they propose that the former involves a greater violation of integrity than the latter. Viewed in the light of the intention of the intervention, tail docking would constitute a greater violation of integrity because this operation is usually conducted only for aesthetic purposes, in contrast to a leg amputation, which is usually carried out for veterinary reasons.

[72] Cf. "To say of someone that he has violated the integrity of an animal does not so much say that someone has the wrong attitude towards animals, but first and foremost that he has performed an intervention in the life of an animal with morally problematic consequences" (De Vries, 2006, 481). And likewise: "In this sense, the concept of animal or phenotypical integrity may be called a consequentialist concept, because arguments that are couched in its terms base the (prima facie) rightness or wrongness of an action, in this case a genetic modification, on the consequences of the action for the resulting animal" (De Vries, 2009, 97).

In summary, integrity can also be interpreted as a relative concept, and thus degrees in the status of integrity are possible.[73] Ways of defining this gradation will be discussed in the section 'Integrity as a practical criterion'.

Integrity of an individual, breeding line or breed, and species

Earlier, it was mentioned that animal integrity came to play a role in the debate on genetic modification because it provided instruments for discussing the moral unease that people felt about this intervention (Macnaghten, 2004). The question that arises now is to what extent it is possible to speak of the integrity of a breeding line, breed, or species. This question will be addressed using an example.

The Green Fluorescent Protein (GFP) gene derived from the jellyfish *Aequorea victoria* and the luciferase gene derived from the firefly *Photinus pyralis* are widely used in laboratory animal research. Inserting these genes into mice or zebrafish, for example, allows the activity of physiological processes to be monitored. This is because the relevant tissues light up when exposed to light (GFP) or luciferin.

Does the insertion of such genes in, say, a zebrafish line constitute an integrity violation? Before this question can be answered, we first need to establish *whose or what's* integrity is involved. Is it the integrity of an individual member of this zebrafish line? Or is it the integrity of the entire zebrafish line? Let us start with the individual fish. It should be noted that genetic modification takes place at very early embryonic stages. At these stages, there is not yet a body in the sense of a division into and connection between organs. So, there is no violation of any physical boundaries in the strict sense (the narrow interpretation of integrity). However, we can speak of a violation of integrity in a broad sense *if the introduction of such genes leads to a loss of certain abilities later in life*. The reference for what counts as a 'loss of abilities' is the set of abilities the animal would have had if the animal were not genetically modified. However, in the case of the zebrafish carrying GFP and/or the luciferase gene, it is hard to imagine what abilities are lost when these fish are kept under laboratory conditions.[74] This involves the problem that consequences at the phenotypic level are not always easy to predict in advance (CBD, 2002).

However, a difficulty now arises: a distinction can be made between the embryos that undergo an intervention and the future offspring of these GM embryos. We mentioned the integrity violation of the embryos that are subject to an intervention. But what of their possible future offspring? Has their integrity also been violated?

There are at least three options. The first is to take the wild-type animal with its abilities as a reference (as according to Verhoog, 1992; 2003; 2007). In that case, the integrity of the offspring of genetically modified zebrafish is also violated. The problem with this option is that it is not (always) clear what 'the' wild-type animal is – as many laboratory animals have been domesticated over decades and millennia – and why that animal possesses the relevant set of abilities.

The second option is to take the breeding line or breed as developed so far as a reference. In that case, the effects in the current generation of biotechnical interventions that took place in previous generations do not 'count' as integrity violations of the current generation. However, a genetic modification in the current generation, if it had negative consequences on the set of abilities, would entail an integrity violation of the animals in the current generation. We are not aware of any author in the literature defending this option. In any case, the arbitrary reference point is an issue with this option.

The third option is to consider the animals' overall body structure and behaviour. This is particularly relevant when abilities are removed, as sometimes in the case of a gene knockout. The description of the whole body structure (including physiology) may still apply to a specific function, however this function is frustrated by the gene knockout. Although Hauskeller (2005) and Kramer and Meijboom (2021) do not use the concept of integrity, but rather the notion of *telos*, this is nevertheless the thrust of their argument.

Both when the wild-type animal is taken as the starting point, and if the overall body structure and behaviour of the animal is taken as the starting point, one can speak of 'breeding involving an integrity violation', being a variation on the well-known expression 'GM animals with a harmful phenotype' (in Dutch 'fok met ongerief').

[73] Several authors have argued that integrity 'evokes' the norm 'respect for integrity' (Vorstenbosch, 1993; Brom, 1997, 135; Hauskeller, 2016). Indeed, if integrity is interpreted as an absolute concept (as according to Bovenkerk, Brom and Van Den Bergh, 2002), then no human intervention can add to it, but only detract from it (Hauskeller, 2016, 29-30). The only appropriate response for humans would seem to be respect for integrity. However, if integrity is understood to be something that can be present to a greater or lesser extent, then respect alone need not be a morally appropriate response. *Benevolence* could also be an appropriate response, for example, namely if the animal's integrity has already been violated. Benevolence to the animal then involves restoring its integrity partially or fully. For example, a dog that is missing a leg due to whatever cause, and is fitted with a prosthesis, can be said to regain some of its integrity.

[74] This could *potentially* be different for zebrafish that are subsequently released into the wild. It could be that such zebrafish have problems reproducing because the light they emit scares off other zebrafish. In that case, the integrity of the zebrafish has been violated.

The aforementioned considered the integrity of individual members of a species. Is it also possible to speak of the integrity of a breeding line, breed, or species? Neither the narrow nor the broad interpretation of integrity seem to suggest that a breeding line, breed or species could have integrity. In both cases, integrity is based on physicality; the broad interpretation of integrity only takes this a little further by arguing that this physicality provides for certain abilities in animals and that morally speaking, the issue lies in taking away these abilities. However, breeding lines, breeds and species have no body, and consequently it is difficult to speak of the integrity of a breeding line, breed or species.[75]

To summarise, some breeding lines involve an inherent integrity violation, i.e. 'breeding involving an integrity violation'. This will more likely be the case if the wild-type animal is taken as the starting point than if the final body structure of the resulting animals is taken as the starting point. So, because of the link between animal abilities and physicality, and the non-physicality of species (only individuals of the species have a body) it is better not to speak of integrity violations of a breeding line, breed or species.

Integrity as a practical criterion

The concept of integrity was used as a review criterion by the Committee on Animal Biotechnology (CBD), which applied the concept for a period of some ten years to assess animal experiments involving the genetic modification of animals (Brom et al., 1996). The CBD went on to make a further distinction between genotypic and phenotypic integrity violations in the genetic modification of animals. The CBD considered "any alteration made to the genome of the animal as a violation of genotypic integrity", but went on to describe the degree of the violation of integrity by considering the phenotypic consequences of the biotechnological interventions (CBD, 2002, 11). In practice, this meant, for example, that "with a relatively large construct, a particular genetic modification can be conditioned or applied so that it is only expressed in one particular organ or tissue. Often, this in fact makes it possible to limit the violation of integrity at the phenotypic level". In other words, the introduction of a larger piece of DNA did not necessarily count as a greater violation of integrity, as this depended on the final phenotype and therefore the violation of phenotypic integrity.

[75] Rolston (2002) takes a somewhat complicated position, because on the one hand he argues that species can possess integrity, and he sees individual members of a species as parts of the species as a whole, while on the other hand, he accepts that "a species...is not a bounded singular" (Rolston, 2002, 7), which still suggests that species have no body. Discussing this further here would be to exceed the scope of this report.

The latter violation mainly involved aberrations in "biological characteristics, such as appearance, behaviour and self-reliance" (CBD, 2008, 22), with the norm for behaviour (and appearance) being dictated by whether or not it is "species-specific" (CBD, 2008, 13).[76]

The example of the CBD shows that, in practice, it is not always easy to define when and to what degree there is a violation of integrity of an animal:

> A recurring difficulty for the CBD is that, in many cases, the violation of health, welfare and integrity of GM animals cannot be reliably measured, and therefore cannot be adequately and scientifically substantiated. Even if these animals were available to researchers with the required expertise and facilities, there is still a lack of validated scientific tools to carry out such research according to a broadly endorsed method. However, GM often involves animals that are not available for such research, or even animals whose phenotype is not yet known. So, for the meantime, the CBD can only evaluate alleged violations of welfare and integrity based on the information and expertise available internally and to the best of its knowledge. The development of valid measurement instruments to this end should therefore receive full attention (CBD, 2002, 13).

The CBD addresses the point that 'valid measurement instruments' are required to assess violations of integrity. Conceivably, and analogous to the distress assessment used in Europe, violations of integrity could be assessed and classified by a working group formed for that purpose.

A question that arises here, is to what extent we can distinguish between degrees of the severity of the violation of integrity? The following in any case applies: If integrity is narrowly interpretated as physical wholeness, it is difficult to imagine how various interventions – such as a vasectomy, toe clipping, or taking a blood sample – could have varying degrees of severity. In all these cases, the physical boundaries are crossed. The gradation will then mainly relate to the frequency of the interventions that violate integrity. Brom (1997) suggests a variation on this: the integrity of a vasectomised male mouse is violated, but is violated again by toe clipping.

[76] The CBD (2008) did not make entirely clear whether 'species-specific' concerned only 'behaviour', or whether it involved 'appearance' too. If 'species-specific' does not concern 'appearance', then the question is, what norm was applied to this characteristic? If it does also involve 'appearance', then questions arise regarding the consistency of the CBD's assessment method. For example, the CBD argued that the phenotypic integrity of animals (such as mice or zebrafish), in which a gene construct from a jellyfish or firefly has been inserted, is not or minimally violated. However, it is difficult to maintain that a zebrafish, that lights up like a firefly on exposure to a certain substance, has a species-specific appearance.

The broad interpretation of integrity is different. The frequency of interventions also plays a role here, but in addition, the severity of integrity violation is mainly related to the degree to which a specific animal ability is affected, and the number of abilities affected. In a vasectomy, for example, the reproductive apparatus as a system is violated in its integrity, loses its reproductive function, and consequently also affects the integrity of the animal as a whole. However, an intervention is conceivable that does not completely eliminate the reproductive function, but instead only reduces fertility. This therefore involves a more limited violation of integrity.

A classification of integrity violations could conceivably be developed with mild, moderate and severe degrees of violation. As with the distress assessment, this will not be a quantitative exercise, but rather a qualitative estimation which preferably will also involve the input of ethologists. Severe integrity violations then form the key focus area for prioritising the replacement of animal experiments.

Summary

It has been argued above that the degree of integrity violation of laboratory animals can be used as a criterion to prioritise the replacement of animal experiments. At the heart of integrity lies the relationship between parts of a given whole: in the case of animals, the relationship between body parts and the animal as a whole, which in itself in turn relates to the whole of its environment. A narrow and a broad interpretation of integrity were discussed, with the former referring to the crossing of the animal's physical boundaries, and the latter referring to any intervention that leads to a loss of species-specific abilities.

The broad interpretation of integrity raises the question of what reference point to use in the case of offspring that are born to animals whose integrity has been violated through genetic modification. Is these offspring's integrity also violated? Both when the wild-type animal is taken as the starting point, and if the overall body structure and behaviour of the animal is taken as the starting point, one can indeed speak of integrity violation of the offspring. This involves a variation on the expression 'GM animals with a harmful phenotype'' (in Dutch 'fok met ongerief'): 'breeding involving an integrity violation'.

Regarding degrees of integrity violation, in the narrow interpretation (intervention in an animal body), only the number of interventions in a single animal can be distinguished. The frequency also plays a role in the broad interpretation, but a further gradation of integrity violation can be made based on the fact that more or less species-specific abilities can be taken away from an animal to a greater or lesser extent.
The broad interpretation of integrity probably does the most justice to people's sense of moral unease concerning, say, nude mice or vasectomised male mice. However, it is not easy to operationalise the broad interpretation of integrity, because this requires a lot of knowledge of the species-specific abilities of animals and whether and when these abilities are affected. A working group will preferably be formed to work on a classification in terms of mild, moderate and severe degrees of integrity violation, linked to the frequency of interventions, the degree to which a specific animal ability is affected, and the number of abilities affected. Animal models and animal experiments with severe forms of integrity violation will then form the key focus area for prioritising the replacement of animal experiments (see also 'Example of integrity: nude mice').

Example of integrity

NUDE MICE

Nude mice lack a thymus and fur. They are the result of a natural mutation. The absence of a thymus means that the mice possess only very few T cells and so are immunodeficient. Nude mice are widely used in oncological animal models (Boven and Winograd, 2018).

Although nude mice did not originate as a result of deliberate human intervention, they can still be said to suffer an integrity violation. The animals are not 'whole', in the sense that they lack fur and a thymus, and so are poorly adapted to changing environmental conditions: nude mice are more likely to become undercooled and ill. Consequently, these animals cannot live independently, but require a highly controlled microbiological and climatic laboratory environment. This means the degree of integrity violation in nude mice is greater than, for example, in mice that have undergone a thymectomy. Nude mice do not only lack a thymus, but they also do not have fur and thus a relative ability to adapt to changing climatic conditions.

A concrete example of a potential key focus area for prioritising the replacement of animal experiments is the Red Fluorescent Protein (RFP) Nude Mouse. This is a nude mouse in which the gene for Red Fluorescent Protein obtained from the cnidarian Discosoma striata *was inserted into the germ line. Cells from a human cancer cell line (in which the gene for Green Fluorescent Protein obtained from the jellyfish* Aequorea victoria *is inserted) are then orthotopically transplanted into these mice. This allows researchers to study the microenvironment of the tumour.*

According to a narrow interpretation of integrity, it is particularly the orthotopic implantation that forms an issue here. This is also the case for the broad interpretation of integrity, because it is likely that certain abilities will be affected by the developing tumour, for example with regard to locomotion. However, the state of nudeness and the lack of a thymus then also form a problem (as discussed above). It is more difficult to say whether RFP and GFP violate integrity. This will require further study.

Assessment using a framework of criteria

The aim of this study is to identify criteria for prioritising the replacement of animal experiments and animal models. It is important to note here that this issue can be discussed both at the level of the assessment of an animal experiment based on the four primary criteria, and at the level of the empirical assessment of how a specific type of animal experiment scores for a criterion or subcriterion. Tools developed to facilitate the latter form of assessment can be found in the literature (see, for example, the section on 'Tools for determining the validity of animal models' for the case of translatability). When choosing to use a quantitative assessment tool like this, it should be kept in mind that it may well create a semblance of objectivity, as the choices made will be integrated into the developed scoring system (Grimm, Olsson and Sandøe, 2019).

Based on the various criteria, an assessment can be carried out into whether a specific type of animal experiment should be placed higher or lower on the hypothetical list of animal experiments that should be replaced as soon as possible. It is worth noting here that there may be both synergisms and conflicts between the various criteria. An example of the former is the work on semi-natural habitats for laboratory animals from the point of view of translatability, which also benefits the welfare of laboratory animals. An example of a conflict between criteria is a thymectomy versus using a line of nude mice. The use of a nude mouse line involves a greater violation of integrity because all individuals in that breeding line will exhibit the phenotype of lacking a thymus, as opposed to the few mice that will undergo a thymectomy. A thymectomy, however, brings inherent (albeit moderate) distress. The existence of synergisms or conflicts, and how to deal with these, will have to be considered on a case-by-case basis.

Conclusion

The ultimate goal of this report is to make a first contribution to establishing criteria for prioritising the development of non-animal alternatives to animal experiments and animal models. This report reveals that the degree of translatability of the research, the degree of relevance of the goal of an animal experiment, the degree of distress caused to the animals, and the degree of violation of their integrity provide adequate starting points to be used as criteria for prioritising the replacement of animal experiments.

Translatability concerns determining the reproducibility and generalisability of animal models used in animal experiments. For example, there is debate in the scientific literature on the question of the generalisability, and hence translatability, of the forced swim test in depression research: does this test really say something about depression in humans, or does it merely say something about coping strategies of rodents – and perhaps humans – in stressful contexts? Animal experiments that use animal models with a low degree of translatability within a given field of research have first priority for replacement, also depending on the assessments of other criteria, such as the relevance of the goal of an animal experiment.

Two factors play a role in determining the relevance of the goal of an animal experiment. The first is the type of goal pursued by the research, such as research to develop or improve cosmetics, food, health, chemicals, or otherwise. The second factor is the variant within a type of research, such as health research to cure or alleviate migraine or cancer. In general, a distinction by type of research goal is a useful criterion for prioritising the replacement of animal experiments. Research goals can thereby be classified as basic, serious, or peripheral, where animal experiments that serve peripheral goals have priority for replacement. Examples of the latter are safety tests on chemicals used in the plastics industry and research into post-weaning diarrhoea in piglets.

The third criterion for measuring the degree of distress in laboratory animals builds on existing legislation that covers distress evaluations and assessments of laboratory animal research. The degree of distress caused to an animal concerns the pain, suffering, fear or lasting harm that an individual animal may suffer because of an intervention. Dutch legislation follows the European legislation in classifying the degree of distress in terms of non-recovery, mild, moderate and severe. Animal experiments that cause severe and prolonged distress have first priority for replacement, followed by animal experiments involving any severe distress. An example of the latter are *batch potency* tests, where the efficacy of each batch of vaccines needs to be tested.

The criterion for the degree of violation of integrity refers to any physical intervention (narrow interpretation of integrity), or any intervention that results in a loss of species-specific abilities (broad interpretation of integrity). Animal experiments that constitute a severe violation of integrity have the highest priority for replacement. Whether an intervention entails a mild, moderate or severe violation of integrity will depend mainly on the frequency of the integrity violations (in the case of the narrow interpretation of integrity), and will also depend on the extent to which animal abilities are affected and the number of abilities affected (in the case of the broad interpretation of integrity). Analogous to the classification of distress in animal experiments, this could be further elaborated, for example by a working group.

The animal species used, the degree of proportionality of the animal experiments, and the numbers of animals or procedures involved are considered to be unusable as criteria for the replacement of animal experiments. While the development phase of alternatives for animal experiments is strategically important, it is not relevant as an argument for replacing animal experiments that are ethically most problematic.

This report concludes that the degree of translatability of the research, the degree of relevance of the goal of an animal experiment, and the degree of distress and violation of integrity caused to the laboratory animals, provide adequate starting points for prioritising the replacement of animal experiments. Based on these criteria, a list of priorities can be conceived based on types of animal experiments and animal models that should be replaced as soon as possible. Examples of types of animal experiments that score poorly on at least one criterion include the forced swim test, research on post-weaning diarrhoea in pigs, batch testing of Clostridium vaccines, and the use of genetically modified nude mice.

Recommendations

Based on the interviews and the literature review, we have the following recommendations for the various stakeholders:

- Laboratory animal experts, ethologists, and scientists with related expertise should form a working group in which a classification – e.g. in terms of mild, moderate and severe – of integrity violations is developed, analogous to the classification of distress. It is important to do this at the European level to ensure effective harmonisation.
- Researchers who work with animal models (both in academia and the industry) and philosophers of science should form a working group in which a classification – e.g. in terms of low, medium and high – of translatability is developed. Existing tools for establishing translatability can provide a starting point here. Again, it is important to do this in the European context.
- Research boards and funding bodies should apply the four criteria (the degree of translatability of the research, the degree of relevance of the goal of an animal experiment, the degree of distress caused to the animals, and the degree of violation of their integrity) as a framework to encourage non-animal technologies there where the ethical priority is highest. Ideally, this should coincide with support for non-animal alternatives that are already at an advanced stage of development (see also 'The stage of development of alternatives'). Such instruments may include funding (e.g. through NWO), public-private partnerships and PR campaigns (for more examples, see Borrás and Edquist, 2013).
- Animal experiment policy advisers and animal welfare representatives should work together with research boards and researchers, biotechnicians and animal carers working in specific research fields to identify animal experiments and animal models where non-animal alternatives are urgently needed according to the four criteria. This will not be easy, as was revealed during the production of this report. Researchers often have a certain interest in using specific animal models and animal experiments, and the public debate is highly polarised. Moreover, animal experiments are used in very different contexts, which makes it difficult to establish a valid point of departure. Promoting innovation 'from within', that is, by the practisers of animal experiments themselves, is possibly the most important in this matter. The development of a 'prioritisation tool' could be useful in this regard, as science boards and researchers, biotechnicians and animal caretakers within a specific field would be able to deploy it themselves. In doing so, it will be important to highlight the value of such a tool both financially (animal experiments are not cheap), strategically ('licence to science'), scientifically (a poorly translatable experiment = substandard science) and ethically (care of animals).

Bibliography

Alliance for Human Relevant Science. 2020.
'Accelerating the Growth of Human Relevant Life Sciences in the United Kingdom: A White Paper by the Alliance for Human Relevant Science'.
https://www.humanrelevantscience.org/wp-content/uploads/Accelerating-the-Growth-of-Human-Relevant-Sciences-in-the-UK_2020-final.pdf

Animals (Scientific Procedures) Act 1986. 2013.

Animals Today. 2015.
'Labradors Universiteit Maastricht door oog van de naald'. 2015.
https://www.animalstoday.nl/labradors-universiteit-maastricht-oog-naald/

Bailoo, Jeremy D., Eimear Murphy, Maria Boada-Saña, Justin A. Varholick, Sara Hintze, Caroline Baussière, Kerstin C. Hahn, e.a. 2018. 'Effects of Cage Enrichment on Behavior, Welfare and Outcome Variability in Female Mice'. *Frontiers in Behavioral Neuroscience* 12.
https://www.frontiersin.org/article/10.3389/fnbeh.2018.00232

Balzer, Philipp, Klaus Peter Rippe, en Peter Schaber. 2000. 'Two Concepts of Dignity for Humans and Non-Human Organisms in the Context of Genetic Engineering'. *Journal of Agricultural and Environmental Ethics* 13 (1): 7–27.
https://doi.org/10.1023/A:1009536230634

Bauer, Thomas R., Yu-Chen Gu, Kate E. Creevy, Laura M. Tuschong, Lisa Embree, Steven M. Holland, Robert A. Sokolic, en Dennis D. Hickstein. 2004. 'Leukocyte Adhesion Deficiency in Children and Irish Setter Dogs'. *Pediatric Research* 55 (3): 363–67.
https://doi.org/10.1203/01.PDR.0000111287.74989.1B

Belzung, Catherine, en Maël Lemoine. 2011. 'Criteria of validity for animal models of psychiatric disorders: focus on anxiety disorders and depression'. *Biology of Mood & Anxiety Disorders* 1 (1): 9.
https://doi.org/10.1186/2045-5380-1-9

Berger Eriksen, Tommy, Thomas Fraser, Håkon Gregersen, Mads Kristiansen, Marianne Olufsen, Binoy Rajan, Raul Ramirez, Jareeporn Ruangsri, Øyvind Røn, en Michel Guarjardo Sarmiento. 2011. 'Should fin clipping be used as a method for identification of fish?' Ås: Norecopa. https://norecopa.no/media/7290/fin-clipping-fish.pdf

Bertzbach, Luca D., Daria Vladimirova, Kristina Dietert, Azza Abdelgawad, Achim D. Gruber, Nikolaus Osterrieder, en Jakob Trimpert. 2021. 'SARS-CoV-2 Infection of Chinese Hamsters (Cricetulus Griseus) Reproduces COVID-19 Pneumonia in a Well-Established Small Animal Model'. *Transboundary and Emerging Diseases* 68 (3): 1075–79.
https://doi.org/10.1111/tbed.13837

Borrás, Susana, en Charles Edquist. 2013. 'The choice of innovation policy instruments'. *Technological Forecasting and Social Change* 80 (8): 1513–22.
https://doi.org/10.1016/j.techfore.2013.03.002

Boven, Epie, en Benjamin Winograd. 2018. *The Nude Mouse in Oncology Research*. Boca Ration, FL: CRC Press.

Bovenkerk, Bernice, Frans W.A. Brom, en Babs J. Van Den Bergh. 2002. 'Brave New Birds: The Use of "Animal Integrity" in Animal Ethics'. *Hastings Center Report* 32 (1): 16–22.
https://doi.org/10.2307/3528292

Brom, Frans W.A. 1997. *Onherstelbaar verbeterd: Biotechnologie bij dieren als een moreel probleem*. Assen: Van Gorcum.

Brom, Frans W.A., M.T. Hilhorst, R.H.J. Ter Meulen, en J.M.G. Vorstenbosch. 1996. 'Het toetsen van biotechnologische handelingen bij dieren. Rapport van een commissie van externe deskundigen ten behoeve van de Commissie Biotechnologie bij Dieren'. Utrecht.
http://docplayer.nl/18577912-Het-toetsen-van-biotechnologische-handelingen-bij-dieren.html
Brønstad, Aurora, Christian E Newcomer, Thierry Decelle, Jeffrey I Everitt, Javier Guillen, en Kathy Laber. 2016. 'Current Concepts of Harm–Benefit Analysis of Animal Experiments – Report from the AALAS–FELASA Working Group on Harm–Benefit Analysis – Part 1'. *Laboratory Animals* 50 (1_suppl): 1–20. https://doi.org/10.1177/0023677216642398

Broom, D M. 1991. 'Animal welfare: concepts and measurement'. *Journal of Animal Science* 69 (10): 4167–75.

Broom. 1993. 'A usable definition of animal welfare'. *Journal of Agricultural and Environmental Ethics* 6 (Supplement 2): 15–25.

Cameron, Daniel, Michael Clemence, en Kyra Xypolia. 2018. 'Public attitudes to animal research in 2018'. London: Ipsos Mori.

Campbell, Donald T., en Donald W. Fiske. 1959. 'CONVERGENT AND DISCRIMINANT VALIDATION BY THE MULTITRAIT-MULTIMETHOD MATRIX'. *Psychological Bulletin* 56 (2).

Campbell, Joy M., Joe D. Crenshaw, en Javier Polo. 2013. 'The biological stress of early weaned piglets'. *Journal of Animal Science and Biotechnology* 4 (1): 19. https://doi.org/10.1186/2049-1891-4-19

CBD. 2002. 'Jaarverslag 2001'. Utrecht: Commissie Biotechnologie bij Dieren.

CBD. 2008. 'Jaarverslag 2007'. Utrecht: Commissie Biotechnologie bij Dieren.

CBG. 2002. 'Integriteit van dieren: Bouwsteen of Struikelblok? Verslag van de eerste discussie-bijeenkomst over biotechnologie bij dieren. Gehouden op 21 november 2002 te Rotterdam'. Utrecht: Centrum voor Bio-ethiek en Gezondheidsrecht. https://edepot.wur.nl/118392

CCD. 2019. 'Ethisch toetsingskader voor proefdiergebruik. Praktische handreiking voor Dierexperimentencommissies'. Centrale Commissie Dierproeven. https://www.centralecommissiedierproeven.nl/documenten/formulieren/16/6/6/ethisch-toetsingskader

CCD. 2021. 'Niet-technische samenvatting 202011508: Opkweek van Europese paling van ei tot glasaal'. Centrale Commissie Dierproeven. https://www.centralecommissiedierproeven.nl/onderwerpen/niet-technische-samenvatting/onderzoek-behoud-diersoort/nts-202011508-europese-paling-glasaal-voortplanting-larven-larvale-voeding-kopie

COGEM. 2018. 'CRISPR & Het Dier: Implicaties van genome editing voor maatschappij en beleid, COGEM signalering CGM/180501-01'. Bilthoven: Commissie Genetische Modificatie.

Commons, Kathryn G., Aram B. Cholanians, Jessica A. Babb, en Daniel G. Ehlinger. 2017. 'The Rodent Forced Swim Test Measures Stress-Coping Strategy, Not Depression-like Behavior'. *ACS Chemical Neuroscience* 8 (5): 955–60. https://doi.org/10.1021/acschemneuro.7b00042

Comstock, Gary. 1992. 'The Moral Irrelevance of Autonomy'. *Between the Species*, 15–27.

Cook, David, Dearg Brown, Robert Alexander, Ruth March, Paul Morgan, Gemma Satterthwaite, en Menelas N. Pangalos. 2014. 'Lessons Learned from the Fate of AstraZeneca's Drug Pipeline: A Five-Dimensional Framework'. *Nature Reviews Drug Discovery* 13 (6): 419–31. https://doi.org/10.1038/nrd4309

Cooper, David E. 1998. 'Intervention, humility and animal integrity'. In *Animal Biotechnology and Ethics*, door Allen Holland en Andrew Johnson, 145–55. Dordrecht: Springer Science + Business Media.

Cooper, Jane, en Maggy Jennings. 2008. 'Advancing animal welfare and the 3Rs in the batch testing of veterinary vaccines'. Horsham, UK: RSPCA.

Cruelty Free Europe. 2022. 'Replace Animal Tests: the RAT list: 10 animal tests that should be replaced'. 2022. https://www.crueltyfreeeurope.org/RATlist

De Kloet, E. R., en M. L. Molendijk. 2016. 'Coping with the Forced Swim Stressor: Towards Understanding an Adaptive Mechanism'. *Neural Plasticity* 2016: 6503162. https://doi.org/10.1155/2016/6503162

De Vries, Rob. 2004. 'Niet alleen ongerief: Waarom DEC's ook zouden moeten toetsen op aantasting van de integriteit van proefdieren'. In *DEC's in discussie: De beoordeling van dierproeven in Nederland*, 61–68. Budel: Damon.

De Vries. 2006. 'Genetic Engineering and the Integrity of Animals'. *Journal of Agricultural and Environmental Ethics* 19 (5): 469–93. https://doi.org/10.1007/s10806-006-9004-y

De Vries. 2009. 'Genetic engineering and our duties to laboratory animals: A theoretical and empirical analysis of the concepts of intrinsic value and animal integrity Rob de'. Nijmegen: Radboud Universiteit Nijmegen.

Degeling, Chris, en Jane Johnson. 2013. 'Evaluating Animal Models: Some Taxonomic Worries'. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 38 (2): 91–106. https://doi.org/10.1093/jmp/jht004

DeGrazia, David, en Tom L. Beauchamp. 2020. '*Principles of Animal Research Ethics*'. In Principles of Animal Research Ethics, onder redactie van Tom L. Beauchamp en David DeGrazia, 5–41. Oxford: Oxford University Press.

Deisseroth, Karl. 2015. 'Optogenetics: 10 years of microbial opsins in neuroscience'. *Nature neuroscience* 18 (9): 1213–25. https://doi.org/10.1038/nn.4091

Denayer, Tinneke, Thomas Stöhr, en Maarten Van Roy. 2014. 'Animal Models in Translational Medicine: Validation and Prediction'. *New Horizons in Translational Medicine* 2 (1): 5–11. https://doi.org/10.1016/j.nhtm.2014.08.001

Eggel, Matthias, en Hanno Würbel. 2021. 'Internal Consistency and Compatibility of the 3Rs and 3Vs Principles for Project Evaluation of Animal Research'. *Laboratory Animals* 55 (3): 233–43. https://doi.org/10.1177/0023677220968583

Engster, Daniel. 2006. 'Care Ethics and Animal Welfare'. *Journal of Social Philosophy* 37 (4): 521–36. https://doi.org/10.1111/j.1467-9833.2006.00355.x

Engster. 2019. 'Care Ethics, Dependency, and Vulnerability'. *Ethics and Social Welfare* 13 (2): 100–114. https://doi.org/10.1080/17496535.2018.1533029

Ethik-Studienkommission des Eidgenössischen Volkswirtschaftsdepartements zur Gentechnologie im ausserhumanen Bereich. 1995. 'Bericht der Ethik-Studienkommission des Eidgenössischen Volkswirtschaftsdepartements Gentechnologie im ausserhumanen Bereich'.

European Commission. 2020. 'REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL. 2019 report on the statistics on the use of animals for scientific purposes in the Member States of the European Union in 2015-2017'. Brussels: European Commission.

European Parliament. 2021. 'European Parliament Resolution of 16 September 2021 on Plans and Actions to Accelerate a Transition to Innovation without the Use of Animals in

Research, Regulatory Testing and Education (2021/2784(RSP))'. https://www.europarl.europa.eu/doceo/document/TA-9-2021-0387_EN.html

European Parliament and the Council. 2019. *Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on Cosmetic Products (Recast)*. http://data.europa.eu/eli/reg/2009/1223/2019-08-13/eng

Europees Parlement en de Raad van de Europese Unie. 2010. 'RICHTLIJN 2010/63/EU VAN HET EUROPEES PARLEMENT EN DE RAAD van 22 september 2010 betreffende de bescherming van dieren die voor wetenschappelijke doeleinden worden gebruikt'. *Publicatieblad van de Europese Unie*, 20 oktober 2010.

Eurostat. 2021. 'Glossary:Basic research'. Eurostat: Statistics Explained. 20 juli 2021. https:/ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Basic_research

Expert Working Group for Project Evaluation and Retrospective Assessment. 2011. 'National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes: Working document on Project Evaluation and Retrospective Assessment'. https://ec.europa.eu/environment/chemicals/lab_animals/pdf/Consensus_document.pdf

Fenwick, Nicole, Gilly Griffin, en Clément Gauthier. 2009. 'The welfare of animals used in science: How the "Three Rs" ethic guides improvements'. *The Canadian Veterinary Journal* 50 (5): 523–30.

Ferreira, Guilherme S., Désirée H. Veening-Griffioen, Wouter P. C. Boon, Ellen H. M. Moors, Christine C. Gispen-de Wied, Huub Schellekens, en Peter J. K. van Meer. 2019. 'A standardised framework to identify optimal animal models for efficacy assessment in drug development'. *PLoS ONE* 14 (6): e0218014. https://doi.org/10.1371/journal.pone.0218014

Ferreira, Guilherme S., Désirée H. Veening-Griffioen, Wouter P. C. Boon, Ellen H. M. Moors, en Peter J. K. van Meer. 2020. 'Levelling the Translational Gap for Animal to Human Efficacy Data'. *Animals* 10 (7): 1199. https://doi.org/10.3390/ani10071199

Fraser, D. 1995. 'Science, Values and Animal Welfare: Exploring the "Inextricable Connection"'. *Animal Welfare* 4 (2): 103–17.

Fraser, D. 2008. 'Animal Welfare and the Intensification of Animal Production'. In *The Ethics of Intensification: Agricultural Development and Cultural Change*, door Paul B. Thompson, 2008:167–89. The International Library of Environmental, Agricultural and Food Ethics 16. Dordrecht: Springer Science + Business Media B.V.

Fraser, D., D. M. Weary, E. A. Pajor, en B. N. Milligan. 1997a. 'A scientific conception of animal welfare that reflects ethical concerns'. *Animal Welfare* 6 (3): 187–205.

Fraser, D., D. Weary, E. Pajor, en B. Milligan. 1997b. 'A Scientific Conception of Animal Welfare that Reflects Ethical Concerns'. *Ethics and Animal Welfare*, januari. http://animalstudiesrepository.org/ethawel/1

Freriks, A.A., B.M.J. Van der Meulen, H. Van den Belt, H. Ten Holt, en J. Verstappen. 2005. 'Noodzakelijk kwaad: Evaluatie Wet op de dierproeven'. https://edepot.wur.nl/35442

Garner, Joseph P. 2014. 'The Significance of Meaning: Why Do over 90% of Behavioral Neuroscience Results Fail to Translate to Humans, and What Can We Do to Fix It?' *ILAR Journal* 55 (3): 438–56. https://doi.org/10.1093/ilar/ilu047

Garner, Joseph P., Brianna N. Gaskill, Elin M. Weber, Jamie Ahloy-Dallaire, en Kathleen R. Pritchett-Corning. 2017. 'Introducing Therioepistemology: The Study of How Knowledge Is Gained from Animal Research'. *Lab Animal* 46 (4): 103–13. https://doi.org/10.1038/laban.1224

Geyer, Mark, en Athina Markou. 1995. 'Animal Models of *Psychiatric Disorders*'. *In Psychopharmacology: The Fourth Generation of Progress*, onder redactie van Floyd E. Bloom en David J. Kupfer, 787–98. New York: Raven Press, Ltd.

Gjerris, Mickey, en C. Gamborg. 2010. 'Is there more to life than welfare? How the concept of animal integrity can contribute to discussions of animal ethics'. In *Global food security: ethical and legal challenges. EurSafe 2010, Bilbao, Spain, 16-18 September, 2010*, 372–77. Wageningen: Wageningen Academic Publishers.

Gjerris, Mickey, en Peter Sandøe. 2006. 'Farm animal cloning: The role of the concept of animal integrity in debating and regulating the technology'. In *Ethics and the Politics of Food, 6th EurSafe Congress, 22-24 June 2006, Oslo, Norway*, door Matthias Kaiser en Marianne Lien, 320–24. Wageningen: Wageningen Academic Publishers.

Greek, Ray, en Lisa A. Kramer. 2019. 'The Scientific Problems with Using Non-human Animals to Predict Human Response to Drugs and Disease'. In *Animal Experimentation: Working Towards a Paradigm Change*, onder redactie van Kathrin Herrmann en Kimberley Jayne, 391–416. Leiden: Brill. http://www.jstor.org/stable/10.1163/j.ctvjhzq0f.8

Green, T. C., en D. J. Mellor. 2011. 'Extending ideas about animal welfare assessment to include "quality of life" and related concepts'. *New Zealand Veterinary Journal* 59 (6): 263–71. https://doi.org/10.1080/00480169.2011.610283

Grimm, Herwig. 2014. 'Ethics in Laboratory Animal Science'. In *Comparative Medicine: Anatomy and Physiology*, onder redactie van Erika Jensen-Jarolim, 281–300. Vienna: Springer. https://doi.org/10.1007/978-3-7091-1559-6_15

Grimm, Herwig, I Anna S Olsson, en Peter Sandøe. 2019. 'Harm–benefit analysis – what is the added value? A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research'. *Laboratory Animals* 53 (1): 17–27. https://doi.org/10.1177/0023677218783004

Gross, Alexandra N., S. Helene Richter, A. Katarina J. Engel, en Hanno Würbel. 2012. 'Cage-Induced Stereotypies, Perseveration and the Effects of Environmental Enrichment in Laboratory Mice'. *Behavioural Brain Research* 234 (1): 61–68. https://doi.org/10.1016/j.bbr.2012.06.007

Hau, Jann, en Steven J. Schapiro. 2011. *Handbook of Laboratory Animal Science, Volume II: Animal Models*. Boca Ration, Fl: CRC Press.

Hauskeller, Michael. 2005. 'Telos: The Revival of an Aristotelian Concept in Present Day Ethics'. Inquiry 48 (1): 62–75. https://doi.org/10.1080/00201740510015356

Hauskeller. 2007. 'The Reification of Life'. *Genomics, Society and Policy* 3 (2): 70. https://doi.org/10.1186/1746-5354-3-2-70

Hauskeller. 2016. *Biotechnology and the Integrity of Life: Taking Public Fears Seriously*. London: Routledge.

Haverkamp, Beatrijs, Bernice Bovenkerk, en Marcel F Verweij. 2018. 'A Practice-Oriented Review of Health Concepts'. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 43 (4): 381–401. https://doi.org/10.1093/jmp/jhy011

Hay, Michael, David W. Thomas, John L. Craighead, Celia Economides, en Jesse Rosenthal. 2014. 'Clinical Development Success Rates for Investigational Drugs'. *Nature Biotechnology* 32 (1): 40–51. https://doi.org/10.1038/nbt.2786

Heeger, F.R., en Frans W.A. Brom. 2001. 'Beyond feeling well: Our direct duties towards animals'. In *EurSafe 2011. 'Food Safety, Food Quality and food Ethics'. The Third Congress of the European Society for Agricultural and Food Ethics. 3-5 October 2001. Florence, Italy.*, 243–46. Milan: A&O.

Held, J.R. 1980. 'Muhlbock memorial lecture: Considerations in the provision and characteriza-

tion of animal models'. In *Animal Quality and Models in Biomedical Research, 7th ICLAS Symposium Utrecht 1979*. Stuttgart, Germany: Gustav Fisher Verlag.

Henderson, Valerie C., Jonathan Kimmelman, Dean Fergusson, Jeremy M. Grimshaw, en Dan G. Hackam. 2013. 'Threats to Validity in the Design and Conduct of Preclinical Efficacy Studies: A Systematic Review of Guidelines for In Vivo Animal Experiments'. *PLOS Medicine* 10 (7): e1001489. https://doi.org/10.1371/journal.pmed.1001489

Holdrege, Craig. 2002. 'Seeing the Integrity and Intrinsic Value of Animals: Developing Appreciative Modes of Understanding'. In *Genetic Engineering and the Intrinsic Value and Integrity of Animals and Plants. Proceedings of a Workshop at the Royal Botanical Garden, Edinburgh, UK, 18-21 September 2002*. Llanystumdwy, UK: International Forum for Genetic Engineering.

Holmes, Philip V. 2003. 'Rodent Models of Depression: Reexamining Validity without Anthropomorphic Inference'. *Critical Reviews in Neurobiology* 15 (2): 143–74. https://doi.org/10.1615/critrevneurobiol.v15.i2.30

Home Office. 2021. 'Annual Statistics of Scientific Procedures on Living Animals Great Britain 2020. Presented to Parliament pursuant to section 21(7) and 21A(1) of the Animals (Scientific Procedures) Act 1986 Ordered by the House of Commons to be printed 15 July 2021'. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1002895/annual-statistics-scientific-procedures-living-animals-2020.pdf

Hooijmans, Carlijn R., Marlies Leenaars, en Merel Ritskes-Hoitinga. 2010. 'A Gold Standard Publication Checklist to Improve the Quality of Animal Studies, to Fully Integrate the Three Rs, and to Make Systematic Reviews More Feasible'. *Alternatives to Laboratory Animals* 38 (2): 167–82. https://doi.org/10.1177/026119291003800208

Hursthouse, Rosalind. 2011. 'Virtue Ethics and the Treatment of Animals'. In *The Oxford Handbook of Animal Ethics*, onder redactie van Tom L. Beauchamp en Raymand G. Frey, 119–43. Oxford: Oxford University Press.

Insel, Thomas R. 2007. 'From Animal Models to Model Animals'. *Biological Psychiatry* 62 (12): 1337–39. https://doi.org/10.1016/j.biopsych.2007.10.001

Janssens, Monique. 2020. *Dierenzaken: Blinde vlek van bedrijven*. Management Boek.

Kagan, Shelly. 2016. 'What's Wrong with Speciesism? (Society for Applied Philosophy Annual Lecture 2015)'. *Journal of Applied Philosophy* 33 (1): 1–21. https://doi.org/10.1111/japp.12164

Kagen, sShelly. 2018. 'For Hierarchy in Animal Ethics'. *Journal of Practical Ethics* 6 (1): 1–18.

Kittay, Eva Feder. 2011. 'The Ethics of Care, Dependence, and Disability'. Ratio Juris 24 (1): 49–58. https://doi.org/10.1111/j.1467-9337.2010.00473.x

Knight, Andrew. 2019. 'Critically Evaluating Animal Research'. In *Animal Experimentation: Working Towards a Paradigm Change*, onder redactie van Kathrin Herrmann en Kimberley Jayne, 321–40. Leiden: Brill. http://www.jstor.org/stable/10.1163/j.ctvjhzqof.8

Knight, Sarah, en Louise Barnett. 2008. 'Justifying Attitudes toward Animal Use: A Qualitative Study of People's Views and Beliefs'. *Anthrozoös* 21 (1): 31–42. https://doi.org/10.2752/089279308X274047

Kola, Ismail, en John Landis. 2004. 'Can the Pharmaceutical Industry Reduce Attrition Rates?' *Nature Reviews Drug Discovery* 3 (8): 711–16. https://doi.org/10.1038/nrd1470

Korte, S. Mechiel, Berend Olivier, en Jaap M. Koolhaas. 2007. 'A new animal welfare concept based on allostasis'. *Physiology & Behavior* 92 (3): 422–28. https://doi.org/10.1016/j.physbeh.2006.10.018

Kramer, K., en F. L. B. Meijboom. 2021. 'Using Breeding Technologies to Improve Farm Animal Welfare: What Is the Ethical Relevance of Telos?' *Journal of Agricultural and Environmental Ethics* 34 (1): 2. https://doi.org/10.1007/s10806-021-09843-6

Kristensen, M., A. A. Nierenberg, en S. D. Østergaard. 2018. 'Face and Predictive Validity of the ClockΔ19 Mouse as an Animal Model for Bipolar Disorder: A Systematic Review'. *Molecular Psychiatry* 23 (1): 70–80. https://doi.org/10.1038/mp.2017.192

LaFollette, Hugh, en Niall Shanks. 1996. *Brute Science: Dilemmas of Animal Experimentation*. London: Routledge.

Lammerts van Bueren, E. T., P. C. Struik, M. Tiemens-Hulscher, en E. Jacobsen. 2003. 'Concepts of Intrinsic Value and Integrity of Plants in Organic Plant Breeding and Propagation'. *Crop Science* 43 (6): 1922–29. https://doi.org/10.2135/cropsci2003.1922

Landi, Margaret, Jeffrey Everitt, en B Berridge. 2021. 'Bioethical, Reproducibility, and

Translational Challenges of Animal Models'. *ILAR Journal*, nr. ilaa027 (maart). https://doi.org/10.1093/ilar/ilaa027

Leenaars, Cathalijn H. C., Carien Kouwenaar, Frans R. Stafleu, André Bleich, Merel Ritskes-Hoitinga, Rob B. M. De Vries, en Franck L. B. Meijboom. 2019. 'Animal to Human Translation: A Systematic Scoping Review of Reported Concordance Rates'. *Journal of Translational Medicine* 17 (1): 1–22. https://doi.org/10.1186/s12967-019-1976-2

Lijmbach, Susanne. 1996. 'De actualiteit van Heideggers dierfilosofie'. In *Heidegger en de wereld van het dier*, onder redactie van Bert Blans en Susanne Lijmbach, 63–75. Assen: Van Gorcum & Comp.

Lund, Thomas Bøker, Jesper Lassen, en Peter Sandøe. 2012. 'Public Attitude Formation Regarding Animal Research'. *Anthrozoös* 25 (4): 475–90. https://doi.org/10.2752/175303712X13479798785896

Lund, Thomas Bøker, Morten Raun Mørkbak, Jesper Lassen, en Peter Sandøe. 2014. 'Painful Dilemmas: A Study of the Way the Public's Assessment of Animal Research Balances Costs to Animals against Human Benefits'. *Public Understanding of Science* 23 (4): 428–44. https://doi.org/10.1177/0963662512451402

Lund, Vonne. 2006. 'Natural living—a precondition for animal welfare in organic farming'. *Livestock Science* 100 (2–3): 71–83. https://doi.org/10.1016/j.livprodsci.2005.08.005

Luvisetto, Siro, Ornella Rossetto, Cesare Montecucco, en Flaminia Pavone. 2003. 'Toxicity of Botulinum Neurotoxins in Central Nervous System of Mice'. *Toxicon: Official Journal of the International Society on Toxinology* 41 (4): 475–81. https://doi.org/10.1016/s0041-0101(02)00370-7

Macnaghten, Phil. 2004. 'Animals in Their Nature: A Case Study on Public Attitudes to Animals, Genetic Modification and "Nature"'. *Sociology* 38 (3): 533–51. https://doi.org/10.1177/0038038504043217

Mak, Isabella WY, Nathan Evaniew, en Michelle Ghert. 2014. 'Lost in translation: animal models and clinical trials in cancer treatment'. *American Journal of Translational Research* 6 (2): 114–18.

Mason, G., R. Clubb, N. Latham, en S. Vickery. 2007. 'Why and how should we use environmental enrichment to tackle stereotypic behaviour?' *Applied Animal Behaviour Science* 102 (3–4): 163–88. https://doi.org/10.1016/j.applanim.2006.05.041

Meijboom, Franck L. B., Elzbieta Kostrzewa, en Cathalijn H. C. Leenaars. 2020. 'Joining forces: the need to combine science and ethics to address problems of validity and translation in neuropsychiatry research using animal models'. *Philosophy, Ethics, and Humanities in Medicine* 15 (1): 1. https://doi.org/10.1186/s13010-019-0085-4

Molendijk, Marc L., en E. Ronald De Kloet. 2015. 'Immobility in the Forced Swim Test Is Adaptive and Does Not Reflect Depression'. *Psychoneuroendocrinology* 62 (december): 389–91. https://doi.org/10.1016/j.psyneuen.2015.08.028

Molendijk, Marc L., en E. Ronald De Kloet. 2019. 'Coping with the Forced Swim Stressor: Current State-of-the-Art'. *Behavioural Brain Research* 364 (mei): 1–10. https://doi.org/10.1016/j.bbr.2019.02.005

Molendijk, Marc L., en E. Ronald De Kloet. 2021. 'Forced Swim Stressor: Trends in Usage and Mechanistic Consideration'. *European Journal of Neuroscience* n/a (n/a). https://doi.org/10.1111/ejn.15139

Morrice, Jessica R., Cheryl Y. Gregory-Evans, en Christopher A. Shaw. 2018. 'Animal models of amyotrophic lateral sclerosis: a comparison of model validity'. *Neural Regeneration Research* 13 (12): 2050–54. https://doi.org/10.4103/1673-5374.241445

NCad. 2020. 'Zorgvuldige weging van methoden voor identificatie en genetische karakterisatie'. Den Haag: Nationaal Comité advies dierproevenbeleid.

Nicholson, Pamela, Julia Furrer, Michael Hässig, Christian Strauss, Manfred Heller, Sophie Braga-Lagache, en Joachim Frey. 2019. 'Production of Neutralizing Antibodies against the Secreted Clostridium Chauvoei Toxin A (CctA) upon Blackleg Vaccination'. *Anaerobe* 56 (april): 78–87. https://doi.org/10.1016/j.anaerobe.2019.02.011

Niemi, Steven M. 2021. 'Harm-Benefit Analyses Can Be Harmful'. *ILAR Journal* 60 (3): 341–46. https://doi.org/10.1093/ilar/ilaa016

Nordgren, Anders. 2010. *For Our Children: The Ethics of Animal Experimentation in the Age of Genetic Engineering*. Amsterdam: Rodopi.

NOS. 2015. 'Universiteit Maastricht stopt dierproeven met honden'. 2 juli 2015. https://nos.nl/artikel/2044710-universiteit-maastricht-stopt-dierproeven-met-honden

Novak, Janja, Jeremy D. Bailoo, Luca Melotti, en Hanno Würbel. 2016. 'Effect of Cage-Induced Stereotypies on Measures of Affective State and Recurrent Perseveration in CD-1 and C57BL/6 Mice'. Onder redactie van Kathleen R. Pritchett-Corning. *PLOS ONE* 11 (5): e0153203. https://doi.org/10.1371/journal.pone.0153203

Nussbaum, Martha. 2006. *Een waardig bestaan. Over dierenrechten*. Vertaald door Rogier Van Kappel. Amsterdam: Ambo.

NVWA. 2022. 'Zo doende 2020. Jaaroverzicht dierproeven en proefdieren van de Nederlandse Voedsel- en Warenautoriteit'. Utrecht: Nederlandse Voedsel- en Warenauthoriteit van het Ministerie van Landbouw, Natuur en Voedselkwaliteit.

Olsson, I Anna S, Christine J Nicol, Steven M Niemi, en Peter Sandøe. 2019. 'From Unpleasant to Unbearable—Why and How to Implement an Upper Limit to Pain and Other Forms of Suffering in Research with Animals'. *ILAR Journal* 60 (3): 404–14. https://doi.org/10.1093/ilar/ilz018

Olsson, I. Anna S., en Peter Sandøe. 2021. 'Chapter 1 - Animal Research Ethics'. In *Handbook of Laboratory Animal Science: Essential Principles and Practices*, onder redactie van Jann Hau en Steven J. Schapiro. Boca Raton, Fl: CRC Press.

Olsson, I Anna S, Orsolya Varga, en Peter Sandøe. 2015. 'A Matter of Importance: Considering Benefit in Animal Ethics Review'. *ALTEX - Proceedings* 4 (1): 33–36.

Ormandy, Elisabeth H., en Catherine A. Schuppli. 2014. 'Public Attitudes toward Animal Research: A Review'. *Animals* 4 (3): 391–408. https://doi.org/10.3390/ani4030391

Ortiz, Sara Elizabeth Gavrell. 2004. 'Beyond Welfare: Animal Integrity, Animal Dignity, and Genetic Engineering'. *Ethics and the Environment* 9 (1): 94–120.

PETA. 2020. 'The Research Modernisation Deal'. London: People for the Ethical Treatment of Animals Foundation.

PETA. 2022. 'Victories! PETA Is Ending Near-Drowning Experiments on Animals'. 2022. https://www.peta.org/features/peta-ends-near-drowning-tests-small-animals/

Porsolt, R. D., M. Le Pichon, en M. Jalfre. 1977. 'Depression: A New Animal Model Sensitive to Antidepressant Treatments'. *Nature* 266 (5604): 730–32. https://doi.org/10.1038/266730a0

Pound, Pandora, en Christine J. Nicol. 2018. 'Retrospective Harm Benefit Analysis of Pre-Clinical Animal Research for Six Treatment Interventions'. *PLOS ONE* 13 (3): e0193758. https://doi.org/10.1371/journal.pone.0193758

Pound, Pandora, en Merel Ritskes-Hoitinga. 2018. 'Is It Possible to Overcome Issues of External Validity in Preclinical Animal Research? Why Most Animal Models Are Bound to Fail'. *Journal of Translational Medicine* 16 (304): 1–8. https://doi.org/10.1186/s12967-018-1678-1

Procaccini, Claudio, Veronica De Rosa, Valentina Pucino, Luigi Formisano, en Giuseppe Matarese. 2015. 'Animal models of Multiple Sclerosis'. *European Journal of Pharmacology* 759 (juli): 182–91. https://doi.org/10.1016/j.ejphar.2015.03.042

Ram, Rebecca. 2019. 'Extrapolation of Animal Research Data to Humans: An Analysis of the Evidence'. In Animal Experimentation: Working Towards a Paradigm Change, onder redactie van Kathrin Herrmann en Kimberley Jayne, 341–75. Leiden: Brill. http://www.jstor.org/stable/10.1163/j.ctvjhzqof.8.

RDA. 2016. 'Fokkerij en Voortplantingstechnieken: Anima(l) sana in corpore sano'. Raad voor Dierenaangelegenheden. https://www.rda.nl/publicaties/zienswijzen/2016/02/29/zienswijze-fokkerij-en-voortplantingstechnieken-uitgebreid-verslag-pdf.

RDA. 2018a. 'Denkkader Dierenwelzijn'. Raad voor Dierenaangelegenheden. https://www.rda.nl/publicaties/publicaties/2018/12/21-rda-denkkader/denkkader

RDA. 2018b. 'Dierproeven ten behoeve van de Veehouderij. Over het belang van dierenwelzijn, duurzaamheid en een ketenbenadering'. Raad voor Dierenaangelegenheden. https://www.rda.nl/publicaties/zienswijzen/2018/10/01/rda-zienswijze-dierproeven-ten-behoeve-van-de-veehouderij

Rhouma, Mohamed, John Morris Fairbrother, Francis Beaudry, en Ann Letellier. 2017. 'Post Weaning Diarrhea in Pigs: Risk Factors and Non-Colistin-Based Control Strategies'. *Acta Veterinaria Scandinavica* 59 (1): 31. https://doi.org/10.1186/s13028-017-0299-7

Röcklinsberg, H., C. Gamborg, en M. Gjerris. 2014. 'A Case for Integrity: Gains from Including More than Animal Welfare in Animal Ethics Committee Deliberations'. *Laboratory Animals* 48 (1): 61–71. https://doi.org/10.1177/0023677213514220

Rogers Brambell, F.W., D.S. Barbour, M.B. Barnett, T.K. Ewer, A. Hobson, H. Pitchforth, W.R. Smith, W.H. Thorpe, en F.J.W. Winship. 1965. 'Report of the Technical Committee to Enquire into the Welfare of Animals kept under Intensive Livestock Husbandry Systems. Presented to Parliament by the Secretary of State for Scotland and the Minister of Agriculture, Fisheries and Food by Command of Her Majesty'. London: Her Majesty's Stationery Office.

Rolston, Holmes. 2002. 'What do we mean by the intrinsic value and integrity of plants and animals?' In *Genetic Engineering and the Intrinsic Value and Integrity of Animals and Plants. Proceedings of a Workshop at the Royal Botanical Garden, Edinburgh, UK, 18-21 September 2002*, 5–10. Llanystumdwy, UK: International Forum for Genetic Engineering.

Russell, William Moy Stratton, en Rex Leonard Burch. 1959. *The Principles of Humane Experimental Technique*. London: Methuen.

Rust, Aleksander, Ciara Doran, Rosalyn Hart, Thomas Binz, Paul Stickings, Dorothea Sesardic, Andrew A. Peden, en Bazbek Davletov. 2017. 'A Cell Line for Detection of Botulinum Neurotoxin Type B'. *Frontiers in Pharmacology* 8. https://www.frontiersin.org/article/10.3389/fphar.2017.00796

Rutgers, Bart, en Robert Heeger. 1999. 'Inherent Worth and Respect for Animal Integrity'. In *Recognizing the Intrinsic Value of Animals: Beyond Animal Welfare*, onder redactie van Marcel Dol, Martje Fentener van Vlissingen, Soemini Kasanmoentalib, Thijs Visser, en Hub Zwart, 41–52. Assen: Uitgeverij Van Gorcum.

Sams-Dodd, Frank. 1999. 'Phencyclidine in the Social Interaction Test: An Animal Model of Schizophrenia with Face and Predictive Validity'. *Reviews in the Neurosciences* 10 (1): 59–90. https://doi.org/10.1515/REVNEURO.1999.10.1.59

Sams-Dodd, Frank. 2006. 'Strategies to Optimize the Validity of Disease Models in the Drug Discovery Process'. *Drug Discovery Today* 11 (7–8): 355–63.

Sandgren, Eric P., Robert Streiffer, Jennifer Dykema, Nadia Assad, en Jackson Moberg. 2020. 'Attitudes toward Animals, and How Species and Purpose Affect Animal Research Justifiability, among Undergraduate Students and Faculty'. *PLOS ONE* 15 (5): e0233204. https://doi.org/10.1371/journal.pone.0233204

Sandøe, Peter, Nuno H Franco, Thomas B Lund, Daniel M Weary, en I Anna S Olsson. 2015. 'Harms to Animals – Can We Agree on How Best to Limit Them?' *ALTEX - Proceedings* 4 (1): 28–32.

Schmidt, Kirsten. 2008. 'Würde oder Integrität - Verlangt die gentechnische Veränderung von Tieren neue tierethische Konzepte?' *ALTEX* 25 (4): 313–20.

Schurgers, R.G. 2005. 'Ethiek in dierexperimentencommissies'. P-UB-2005-02. Utrechtse Wetenschapswinkels, voor maatschappijgericht onderzoek. Utrecht: Universiteit Utrecht. https://dspace.library.uu.nl/handle/1874/45043.

Scruton, Roger. 1998. *Animal Rights and Wrongs*. London, UK: Demos.

Sewell, Fiona, Ian Waterson, David Jones, Mark David Tricklebank, en Ian Ragan. 2021. 'Preclinical Screening for Antidepressant Activity – Shifting Focus Away from the Forced Swim Test to the Use of Translational Biomarkers'. *Regulatory Toxicology and Pharmacology* 125 (oktober): 105002. https://doi.org/10.1016/j.yrtph.2021.105002

Singer, Peter. 1993. *Practical Ethics*. 2nd dr. Cambridge: Cambridge University Press.

Singh, Ashutosh, Rahul Soloman Singh, Phulen Sarma, Gitika Batra, Rupa Joshi, Hardeep Kaur, Amit Raj Sharma, Ajay Prakash, en Bikash Medhi. 2020. 'A Comprehensive Review of Animal Models for Coronaviruses: SARS-CoV-2, SARS-CoV, and MERS-CoV'. *Virologica Sinica* 35 (3): 290–304. https://doi.org/10.1007/s12250-020-00252-z

Singh, K.P., N.S. Parihar, en B.N. Tripathi. 1992. 'Pathology and pathogenesis of Clostridium chauvoei infection in guinea-pigs.' *Indian Journal of Animal Sciences* 62: 611–15.
Sjoberg, Espen A. 2017. 'Logical fallacies in animal model research'. *Behavioral and Brain Functions* 13 (1): 3. https://doi.org/10.1186/s12993-017-0121-8

Slattery, David A., en John F. Cryan. 2012. 'Using the Rat Forced Swim Test to Assess Antidepressant-like Activity in Rodents'. *Nature Protocols* 7 (6): 1009–14. https://doi.org/10.1038/nprot.2012.044

Smith, Jane A., en Kenneth M. Boyd, red. 1991. *Lives in the Balance: The Ethics of Using Animals in Biomedical Research. The Report of a Working Party of the Institute of Medical Ethics*. Oxford: Oxford University Press.

Steel, Daniel. 2010. 'A New Approach to Argument by Analogy: Extrapolation and Chain Graphs'. *Philosophy of Science* 77 (5): 1058–69. https://doi.org/10.1086/656543

Swearengen, James R. 2018. 'Choosing the Right Animal Model for Infectious Disease Research'. *Animal Models and Experimental Medicine* 1 (2): 100–108. https://doi.org/10.1002/ame2.12020

Tannenbaum, J. 1991. 'Ethics and animal welfare: the inextricable connection'. *Journal of the American Veterinary Medical Association* 198 (8): 1360–76.

Thomas, David W., Justin Burns, John Audette, Adam Carroll, Corey Dow-Hygelund, en Michael Hay. 2016. 'Clinical Development Success Rates 2006-2015'. https://www.bio.org/sites/default/files/legacy/bioorg/docs/Clinical%20Development%20Success%20Rates%202006-2015%20-%20BIO,%20Biomedtracker,%20Amplion%202016.pdf

Treit, Dallas, Elif Engin, en Kris McEown. 2010. 'Animal Models of Anxiety and Anxiolytic Drug Action'. In *Behavioral Neurobiology of Anxiety and Its Treatment*, onder redactie van Murray B. Stein en Thomas Steckler, 121–60. Current Topics in Behavioral Neurosciences. Berlin, Heidelberg: Springer. https://doi.org/10.1007/7854_2009_17

Tricklebank, Mark, en Joseph Garner. 2012. 'The Possibilities and Limitations of Animal Models for Psychiatric Disorders'. In *Drug Discovery for Psychiatric Disorders*, onder redactie van Z. Rankovic, M. Bingham, E.J. Nestler, en R. Hargreaves, 28:534–56. London: The Royal Society of Chemistry. https://doi.org/10.1039/9781849734943-00534

Trouw. 1995. 'Onderzoekers laten mensenoor op muis groeien'. *Trouw*, 26 oktober 1995, sec. voorpagina. https://www.trouw.nl/gs-b6fe515d

Trunnell, Emily R., en Constança Carvalho. 2021. 'The Forced Swim Test Has Poor Accuracy for Identifying Novel Antidepressants'. *Drug Discovery Today* 26 (12): 2898–2904. https://doi.org/10.1016/j.drudis.2021.08.003

Tweede Kamer der Staten Generaal. 2021. 'Wijziging van de Wet dieren in verband met de uitvoering van de herziene Europese diergezondheidswetgeving. Amendement van het lid Vestering. Ontvangen 21 april 2021'. https://www.parlementairemonitor.nl/9353000/1/j4nvgs5kjg27kof_j9vvij5epmj1ey0/vli-6ae95lkqr/f=/kst353989.pdf

Van den Heuvel, Steven C., Patrick Nullens, en Angela Roothaan, red. 2017. *Theological Ethics and Moral Value Phenomena: The Experience of Values*. 1st edition. London: Routledge.

Van der Staay, F Josef, Saskia S Arndt, en Rebecca E Nordquist. 2009. 'Evaluation of animal models of neurobehavioral disorders'. *Behavioral and Brain Functions* : BBF 5 (februari): 11. https://doi.org/10.1186/1744-9081-5-11

Van der Worp, H. Bart, David W. Howells, Emily S. Sena, Michelle J. Porritt, Sarah Rewell, Victoria O'Collins, en Malcolm R. Macleod. 2010. 'Can Animal Models of Disease Reliably Inform Human Studies?' *PLOS Medicine* 7 (3): e1000245. https://doi.org/10.1371/journal.pmed.1000245

VanDeVeer, Donald. 1995. 'Interspecific Justice'. In *People, Penguins, and Plastic Trees: Basic Issues in Environmental Ethics*, onder redactie van Christine Pierce en Donald VanDeVeer, 85–99. Belmont, CA: Wadsworth Publishing Company.

Verhoog, Henk. 1992. 'The Concept of Intrinsic Value and Transgenic Animals'. *Journal of Agricultural and Environmental Ethics* 5 (2): 147–60. https://doi.org/10.1007/bf01966357

Verhoog, Henk. 2003. 'Naturalness and the genetic modification of animals'. *Trends in Biotechnology* 21 (7): 294–97. https://doi.org/10.1016/S0167-7799(03)00142-2

Verhoog, Henk. 2004. 'Utilisme en de integriteit van het dier'. In *DEC's in discussie: De beoordeling van dierproeven in Nederland*, 51–60. Budel: Damon.

Verhoog, Henk.. 2007. 'The tension between common sense and scientific perception of animals: recent developments in research on animal integrity'. *NJAS: Wageningen Journal of Life Sciences* 54 (4): 361–73. https://doi.org/10.1016/S1573-5214(07)80009-

Vieira de Castro, Ana Catarina, en I. Anna S. Olsson. 2015. 'Does the Goal Justify the Methods? Harm and Benefit in Neuroscience Research Using Animals'. In *Ethical Issues in Behavioral Neuroscience*, onder redactie van Grace Lee, Judy Illes, en Frauke Ohl, 47–78. Current Topics in Behavioral Neurosciences. Berlin, Heidelberg: Springer. https://doi.org/10.1007/7854_2014_319

Visser, M.B.H., en Henk Verhoog. 1999. 'De aard van het beestje: De morele relevantie van natuurlijkheid'. NWO Ethiek en Beleid.

Voelkl, Bernhard, Hanno Würbel, Martin Krzywinski, en Naomi Altman. 2021. 'The Standardization Fallacy'. *Nature Methods* 18 (1): 5–7. https://doi.org/10.1038/s41592-020-01036-9

Vorstenbosch, J. 1993. 'The concept of integrity. Its significance for the ethical discussion on biotechnology and animals'. *Livestock Production Science*, Biotechnology and Animal Welfare, 36 (1): 109–12. https://doi.org/10.1016/0301-6226(93)90144-7

VWA. 2004. 'Zo doende 2003. Jaaroverzicht van de Voedsel en Waren Autoriteit over dierproeven en proefdieren'. Den Haag: Voedsel en Waren Autoriteit.

Wehling, Martin. 2009. 'Assessing the Translatability of Drug Projects: What Needs to Be Scored to Predict Success?' *Nature Reviews. Drug Discovery* 8 (7): 541–46. https://doi.org/10.1038/nrd2898

Wendler, Alexandra, en Martin Wehling. 2012. 'Translatability Scoring in Drug Development: Eight Case Studies'. *Journal of Translational Medicine* 10 (1): 39. https://doi.org/10.1186/1479-5876-10-39

Wendler. 2017. 'Translatability Score Revisited: Differentiation for Distinct Disease Areas'. *Journal of Translational Medicine* 15 (1): 226. https://doi.org/10.1186/s12967-017-1329-y.

Wensley, Madie R, Mike D Tokach, Jason C Woodworth, Robert D Goodband, Jordan T Gebhardt, Joel M DeRouchey, en Denny McKilligan. 2021. 'Maintaining continuity of nutrient intake after weaning. II. Review of post-weaning strategies'. *Translational Animal Science* 5 (1): txab022. https://doi.org/10.1093/tas/txab022
Wet op de dierproeven. 2021. https://wetten.overheid.nl/BWBR0003081/2021-07-01

Willner, Paul. 1984. 'The Validity of Animal Models of Depression'. *Psychopharmacology* 83 (1): 1–16. https://doi.org/10.1007/BF00427414

Willner, Paul. 1991. 'Methods for Assessing the Validity of Animal *Models of Human Psychopathology*'. In Animal Models in Psychiatry, I, door Alan A. Boulton, Glen B. Baker, en Mathew T. Martin-Iverson, 18:1–24. New Jersey: Humana Press. https://doi.org/10.1385/0-89603-198-5:1

Willner, Paul, en P. J. Mitchell. 2002. 'The Validity of Animal Models of Predisposition to Depression'. *Behavioural Pharmacology* 13 (3): 169–88.

Willner, Paul, Richard Muscat, en Mariusz Papp. 1992. 'Chronic Mild Stress-Induced Anhedonia: A Realistic Animal Model of Depression'. *Neuroscience & Biobehavioral Reviews* 16 (4): 525–34. https://doi.org/10.1016/S0149-7634(05)80194-0

Wong, Chi Heem, Kien Wei Siah, en Andrew W Lo. 2019. 'Estimation of clinical trial success rates and related parameters'. *Biostatistics* 20 (2): 273–86. https://doi.org/10.1093/biostatistics/kxx069

Würbel, Hanno. 2017. 'More than 3Rs: The Importance of Scientific Validity for Harm-Benefit Analysis of Animal Research'. *Lab Animal* 46 (4): 164–66. https://doi.org/10.1038/laban.1220

Zeiss, Caroline J. 2015. 'Improving the Predictive Value of Interventional Animal Models Data'. *Drug Discovery Today* 20 (4): 475–82. https://doi.org/10.1016/j.drudis.2014.10.015

Zhai, Chongkai, Mingda Wang, Hea-Jong Chung, Mehedi Hassan, Seungkoo Lee, Hyeon-Jin Kim, en Seong-Tshool Hong. 2021. 'Roborovski hamster (Phodopus roborovskii) strain SH101 as a systemic infection model of SARS-CoV-2'. *Virulence* 12 (1): 2430–42. https://doi.org/10.1080/21505594.2021.1972201

Appendix A

Table 1: *List of persons interviewed*

Person	Organisation	Details
Bas Blaauboer	Central Authority for Scientific Procedures on Animals (CCD), Utrecht University	Interviewed on 29/09/2021
Anonymous	Radboud UMC	Interviewed on 30/09/2021
Jan Langermans	Biomedical Primate Research Centre, Utrecht University	Interviewed on 30/09/2021
Anonymous	Maastricht University	Interviewed on 04/10/2021
Nelleke Verhave	Animal Welfare Body (IvD), Leiden University Medical Centre/ Netherlands Association of Animal Ethics Committees (NVDEC)	Interviewed on 05/10/2021
Harry Emmen	Charles River Laboratories	Interviewed on 08/10/2021
Erik Baltussen	Charles River Laboratories	Interviewed on 08/10/2021
Lennert Schrader	Netherlands Food and Consumer Product Safety Authority	Duo interview on 14/10/2021
Anonymous	Netherlands Food and Consumer Product Safety Authority	Duo interview on 14/10/2021
Saskia Aan	*Proefdiervrij* (Dutch Society for the Replacement of Animal Testing)	Group interview on 07/10/2021
Janneke Hogervorst	People for the Ethical Treatment of Animals (PETA)	Group interview on 07/10/2021
Elly von Jessen	*Dierenbescherming* (Dutch Society for the Protection of Animals)	Group interview on 07/10/2021
Anonymous	Animal Rights	Group interview on 07/10/2021

Annex 4

Texts on fundamental research in the Working document on Project Evaluation and Retrospective Assessment

Sources: National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes. Working document on Project Evaluation and Retrospective Assessment. Brussels, 18-19 September 2013.[77]

Project objectives for basic research (p. 11):
- The current state of knowledge on which project intends to build;
- The way in which the project will help to advance knowledge.

Goals already achieved by previous project(s) (where applicable) should be included: the progress summarised and an indication of which specific objectives should be achieved through this project.

Acknowledgment that in some areas of basic research expanding knowledge can be a suitable objective in its own right – but should, where possible, be linked to dissemination of results (having regard to IP), and potential longer term benefits.

Assessment of the benefits of basic research (p.21 and 22):
- Better understanding of the issue (increased knowledge – acknowledging the importance of filling a knowledge/information gap);
- Wherever possible, "increased knowledge" as the primary benefit should be linked to a more tangible strategic goal, even though any wider benefits may be much further in the future and less predictable; benefits should go beyond "it would be nice to know";
- Scale of improvement (man, animal, environment) (numbers; size and quality – need informed judgement – orphan drugs may be used in a few people but high impact on individuals) and burden to the society of the problem (both on basic/applied research);
- Acknowledgement of basic research driven by hypotheses – evaluation needs to confirm hypothesis is scientifically sound and realistic;
- Dissemination of information, whether positive or negative, is particularly important for basic research to ensure the benefits are realised.

[77] The 'Working document on Project Evaluation and Retrospective Assessment' was compiled by the working group of national authorities responsible for implementing Directive 2010/63/EU. https://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/project_evaluation/nl.pdf.

Key considerations in the assessment of benefits (p.23):

1. Consider the immediate / short term benefits (product; data; outcome).
2. Consider longer term benefits (product; specific long term).
3. Consider the wider impact (increase of knowledge; translational potential).

Contact details
Netherlands National Committee for the protection of animals used for scientific purposes
PO Box 93118 | 2595 AL The Hague
Email: ncad@rvo.nl | Website: ncadierproevenbeleid.nl

April 2023 | Publication no. 23407178