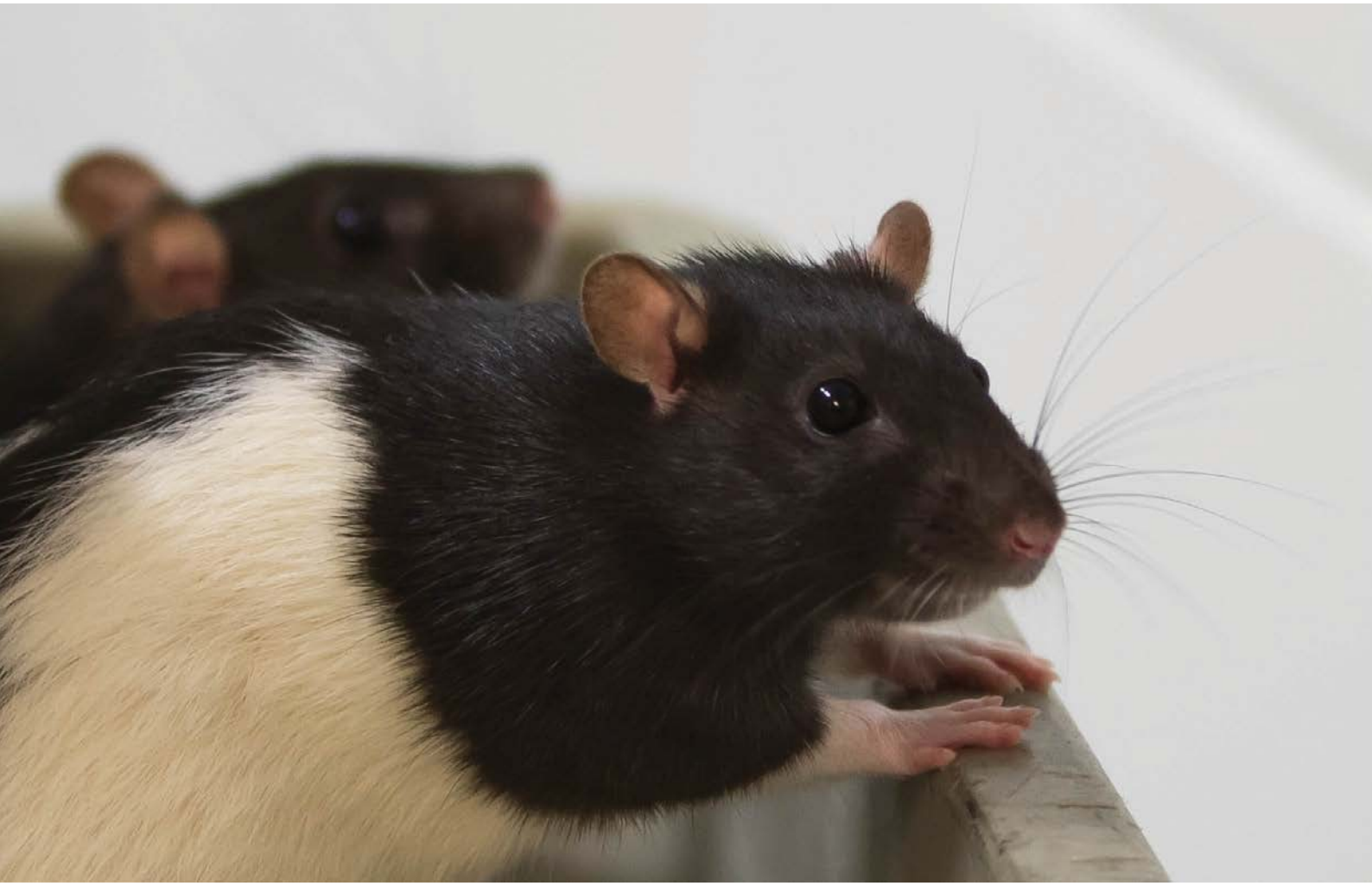
Aus dem Institut der Pharmakologie und Toxikologie des Fachbereichs Veterinärmedizin der Freien Universität Berlin

Refinement on the way towards Replacement of animal experiments:

A retrospective review of the use of Refinement methods in German animal research applications



Inaugural-Dissertation
zur Erlangung des Grades eines
Ph.D. of Biomedical Sciences
an der Freien Universität Berlin

vorgelegt von
Kathrin Herrmann
Tierärztin aus Ulm

Berlin 2019
Journal-Nr.: 4105

Aus dem Institut der Pharmakologie und Toxikologie
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Refinement on the way towards Replacement of animal experiments: A retrospective review of the use of Refinement methods in German animal research applications

Inaugural-Dissertation
zur Erlangung des Grades eines
Ph.D. of Biomedical Sciences
an der
Freien Universität Berlin

vorgelegt von
Kathrin Herrmann
Tierärztin
aus Ulm

Berlin 2019

Journal-Nr.: 4105

Gedruckt mit Genehmigung
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Dekan:	Prof. Dr. Jürgen Zentek
Erste Gutachterin:	Prof. Dr. Heidrun Fink
Zweiter Gutachter:	Prof. Dr. Paul Flecknell
Dritter Gutachter:	Prof. Dr. Michael Erhard

Deskriptoren (nach CAB-Thesaurus): animal welfare, animal testing alternatives, research, Germany, reviews

Tag der Promotion: 08.12.2019

Titelfoto: Cathy Schuppli

This thesis is dedicated to the animals who still suffer in the name of science.

The important thing is not to stop questioning. – Albert Einstein

Table of Contents

1 Introduction

1.1 The 3Rs of animal experimentation

In response to a request from Charles Hume, founder of the Universities Federation for Animal Welfare (UFAW), W.M.S. Russell and R.L. Burch developed the principles of Replacement, Reduction, and Refinement (3Rs) of animal experimentation. Russell and Burch first described the 3Rs six decades ago, in their pioneering book The Principles of Humane Experimental Technique (1959), to remedy “inhumanity” towards non-human animals (hereinafter referred to as animals). With the concept of “inhumanity”, the authors referred to “an objective assessment of the effects of any procedures on the animal subjects” without implying any ethical judgement of the research workers (Chapter 2, Russell and Burch, 1959). They were certain that “inhumanity” could be lessened or eliminated under the three broad categories of Replacement, Reduction, and Refinement of humane technique.

Russell and Burch’s original definitions of the 3Rs are as follows:

- **Replacement** stands for “the substitution for conscious higher animals of insentient material”,

- **Reduction** means “reduction in the number of animals used to obtain information of a given amount and precision” and

- **Refinement** refers to “any decrease in the incidence or severity of inhumane procedures” (Chapter 4, Russell and Burch, 1959).

Their aim was to avoid the use of animals wherever possible and to considerably reform the treatment of the animals still deemed necessary, while significantly improving the study quality by adequate experimental design and statistical analysis (Russell and Burch, 1959).

In recent decades, the 3Rs concept has slowly been given more legal consideration globally. It has been recognized by organizations such as the Council of Europe (1986) and the World Organisation for Animal Health (2018). Today, the principles are generally accepted, and many countries have embedded them into their legislation (Bayne *et al.*, 2015). At this point, the European Directive 2010/63/EU appears to be the most progressive regulation on the protection of animals used for scientific purposes. It not only requires all European Union (EU) Member States to fully implement the 3Rs principles but it is also more far-reaching than other regulations, due to its final goal of “full replacement of procedures on live animals for scientific and educational purposes as soon as it is scientifically possible” (European Parliament, 2010, Recital 10). Moreover, the EU Directive postulates that, for any given experiment, Replacement should be the first priority, followed by Reduction and then Refinement if animal use is deemed unavoidable (European Parliament, 2010, Recital 11). Russell and Burch (1959, Chapter 7) also emphasized the predominant need for Replacement, and, where this is not feasible, the principles of Reduction and Refinement are to be applied: **“**Suppose, for a particular purpose, we cannot use replacing techniques. Suppose it is agreed that we shall be using every device of theory and practice to reduce to a minimum the number of animals we have to employ. It is at this point that refinement starts, and its object is simply to reduce to an absolute minimum the amount of distress imposed on those animals that are still used.”

Despite improved animal protection laws, the number of animals used in science has been on the rise since the 2000s (Taylor and Rego, 2016), yet accompanied by a growth in scrutiny over

both the ethics and the scientific value of animal experimentation. It was estimated that, in the year 2005, more than 127 million vertebrates were used worldwide for scientific purposes (Knight, 2008). With the promotion of easier-to-use genetic modification techniques, such as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), the numbers as well as species of animals used are likely to continue to rise (Bailey, 2019).

One driver for the incorporation of the 3Rs into legislation has notably been societal concern regarding the use and the treatment of animals in scientific research. A recent opinion poll by Clemence and Leaman (2016) found that while the majority of the public accepts the use of animals for medical and scientific purposes, 26% of participants would ban it when animal welfare considerations are not taken into account. 71% of the surveyed public said they accept animal use for medical research if there is no alternative and as long as no needless suffering is caused to the animals (Clemence and Leaman, 2016). In the European Union, citizens can suggest concrete legal changes to the European Commission (EC), by using a participatory democracy tool, the European Citizens' Initiative (European Commission, 2019). The European Citizens' Initiative "Stop Vivisection" showed that a considerable number of EU citizens asked for new legislation that focuses on eliminating animal experimentation and on accelerating human-relevant biomedical and toxicological research. It was signed by almost 1.2 million EU citizens and submitted to the EC in March 2015 (European Commission, 2015). Simultaneously, animal-based research has increasingly been questioned, not only on ethical and animal welfare grounds, but also due to its limited rigor, robustness and translatability (e.g., Begley and Ellis, 2012; Harris, 2017; Pound and Bracken, 2014; Pound and Ritskes-Hoitinga, 2018; Prinz *et al.*, 2011). Notwithstanding extensive preclinical animal testing, less than 12% of drugs entering clinical trials result in an approved medication (Pharmaceutical Research and Manufacturers of America, 2015; 2016). Moreover, it has been estimated that between 51% and 89% of preclinical study results are not reproducible (Freedman *et al.*, 2015; Harthorne and Schachner, 2012). The reproducibility and translatability pitfalls surrounding biomedical (animal) experimentation (e.g., Baker, 2016; Begley and Ioannidis, 2015; Freedman *et al.*, 2017; Harris, 2017; Ritskes-Hoitinga and Pound, 2018) call, according to numerous scientists (e.g., Archibald *et al.*, 2018; Greek and Kramer, 2019; Pound *et al.*, 2004), for a critical appraisal of the scientific viability of animals to model human responses to drugs and diseases. Besides using tools such as systematic reviews, citation and meta-analyses and other retrospective assessments (see also Paper 4, Herrmann, 2019), Greek and Kramer (2019) propose that public investment in animal experiments should be assessed based on its scientific merit. They suggest a peer-reviewed debate evaluating the evidence pro and contra continued animal use that would be conducted in public and judged by unbiased experts (Greek and Kramer, 2019). To overcome the poor translation from animals to humans (discussed e.g., Kramer and Greek, 2018; Pippin, 2012), a growing number of scientists have been promoting a shift in focus towards animal-free, human-biology based and thus human-relevant research (e.g., Archibald *et al.*, 2018; BioMed21 Collaboration, n.d.; Herrmann and Jayne, 2019; Herrmann *et al.*, 2019; Langley at al., 2017).

While replacing live animals will continue to be the ultimate goal (European Parliament, 2010, Recital 10), responsible animal use that encompasses applying all available Refinement methods and that ensures conducting high quality research will remain of utmost importance to justify animal use in science. Russell and Burch recognized the use of Refinement methods not only as the way to limit negative animal welfare implications but also as a prerequisite for successful research. "[B]y now it is widely recognized that the humanest possible treatment of experimental animals, far from being an obstacle, is actually a prerequisite for successful animal experiments" wrote Russell and Burch (1959) in the first chapter of *The Principles of Humane Experimental Technique* where they laid out the scope of the study. When the authors discussed the factors governing scientific progress, they stated under the subsection on "Humanity and

Efficiency": "If we are satisfied that an experiment is maximally humane, we can be quite sure it is the most scientifically valuable one we could perform. This will apply not only to individual experiments but also to whole research programs." (Chapter 8, 1959).

In a 1995 ECVAM Workshop report entitled "The Three Rs: The Way Forward," in which Russell and Burch were among the participants, it was noted that researchers were not sufficiently aware of the concept of Refinement and that they generally did not recognize the impact that the application, or rather the non-application of Refinement, had on their research (Balls *et al.*, 1995). Many more recent literature reviews have also indicated a significant lack of the application of Refinement (as well as of the other 2 Rs; e.g., Bara and Joffe, 2014) in practice (e.g., Carbone and Austin, 2016; Richardson *et al.*, 2005; Pound and Nicol, 2018; Taylor, 2010; Uhlig *et al.*, 2015). Not only does this imply avoidable animal suffering but this lack of Refinement can lead to (chronic) animal stress and distress which can compromise data collected from these animals (see review by Bailey, 2018). Added to this issue is the fact that it is unclear to what extent Refinement is applied in practice, as information about Refinement methods is generally still lacking in published studies describing animal experiments (Bertrand *et al.*, 2018; Carbonne and Austin, 2016; Hair, 2018; Würbel, 2007).

Thus, the first three papers of this publication-based thesis focus on experimental Refinement, the primary aim of which is to reduce to an absolute minimum the amount of pain, distress, suffering or lasting harm caused by the experiment(s). Research into Refinement has notably grown during the past two decades. Nonetheless, the animals can benefit from these efforts only if the newly gained knowledge is applied in practice. In addition to discussing the implementation and use of Refinement, the fourth and final paper of this thesis discusses also the other two Rs, Reduction and Replacement. It highlights tools to retrospectively assess the validity of animal studies, which could lead to a substantial reduction of animal experiments and disease models used and thus to a reduction of the overall number of animals used in experiments. It then reviews recent work towards Replacement, which Russell and Burch (1959) and current EU legislation (European Parliament, 2010, Recital 10) see as the ultimate aim.

1.2 Project aims

Refining both the use and care of animals as well as the quality of research has been widely adopted in the search for means to improve not only animal welfare but reproducibility of results from animal studies. Thus, the primary goal of this project was to review the planned application of Refinement methods during experimentation, namely the use of adequate anesthesia and analgesia protocols, proper pain management and health monitoring, early and thus more humane endpoints ("less-inhumane endpoints" according to Balls, 1999) and acceptable killing methods. In addition, the severity classification of procedures was assessed. After presenting and discussing the findings of what animal experimenters proposed to use in order to refine their experiments (first three papers: Herrmann and Flecknell, 2019; Herrmann and Flecknell, 2018a and 2018b), the final paper (Herrmann, 2019) in this publication-based thesis discusses Refinement in a broader context – by including not only classical Refinement that directly reduces pain, distress and suffering but also Refinement of design, analysis and the reporting of animal studies. Some recent efforts to reduce and replace animal experiments are also briefly discussed.

The aims of this first-of-its-kind retrospective review of animal research proposals, then, were a) to gauge how animal welfare and Refinement research had been implemented into planned research studies by 2010; b) to identify areas where additional improvements of experimental

Refinement are still necessary; c) to provide recommendations for refined practice; and d) to give a brief overview of efforts taken to apply the 3Rs.

1.3 Materials and methods

Since the primary objective of this review was to assess the animal experimenters' intended use of experimental Refinement, a novel approach was taken to assess their planned application of Refinement methods as well as the quality of the chosen Refinement methods. Instead of reviewing published animal research studies, original applications of basic and applied animal research that received licenses were reviewed. The study exclusively assessed proposals originating from Germany, since researchers working in this country are required to provide comprehensive descriptions of all planned procedures, including all Refinement methods, to the competent authorities (Sections 8 and 9, *German Animal Welfare Act* of 1986). German law requires that animals are treated in the most humane way possible while assuring the generation of valid scientific results (Section 9, para. 2, No. 3, *German Animal Welfare Act* of 1986).

Each of the 16 German federal states has its own competent authorities responsible for assessing and licensing animal research applications. For certain scientific research purposes, the relevant authority securely storing the data is entitled to pass on data without consent of persons concerned, if 1. interests requiring protection are not affected through the kind of data, their publication and use, or 2. the public interest in the conduct of the scientific project outweighs the interests of the persons concerned. The data transmission must be approved beforehand by the highest authorities of the federal states (Metschke, R. and Wellbrock, R., 2002, Appendix 3, Section 30, p. 75). In preparation of our study, the Data Protection and Freedom of Information Commissioner of Berlin (Berliner Beauftragter für Datenschutz und Informationsfreiheit) was informed and stated no objections (case number 5612.145). The majority (14 out of 16 federal states) agreed to participate in this study and provided access to basic and applied research proposals that had been licensed. The study was conducted anonymously (i.e. the individual research groups were not identified in the analysis).

For the focus of our study, we chose all applications submitted to the participating competent authorities in 2010 that contained surgical procedures in mice and rats from which the animals were to recover. Choosing such research applications allowed us to review the adequacy of anesthesia, perioperative analgesia, and postoperative pain assessment and pain management. Even though this review was done prior to the implementation of EU Directive 2010/63/EU, the key provisions of the new Directive relative to Refinement had already been included in German legislation in the 1970s and 80s, meaning the use of experimental Refinements was to be discussed in detail in the assessed proposals.

Legal requirements
As of 2010, various prerequisites were required when conducting animal experiments in accordance with the *German Animal Welfare Act (Deutsches Tierschutzgesetz) as amended in 1986:*

- Animal experiments must be reduced to an indispensable extent (Section 9, para. 2, sentence 1, *German Animal Welfare Act*).
- When conducting animal experiments, the current state of scientific knowledge has to be taken into account (Section 9, para. 2, sentence 2, *German Animal Welfare Act*).
- Pain, suffering and harm are allowed to be inflicted upon the animals only thus far as it is indispensable for the striven for purpose (Section 9, para. 2, No. 3, *German Animal Welfare Act*).

- Principal investigators must be adequately qualified (Section 8, para. 3, No. 2, *German Animal Welfare Act*).
- Any person involved in animal research and laboratory animal care must have sufficient professional knowledge (Section 9, para. 1, *German Animal Welfare Act*).
- Principal investigators (who are also the animal research proposal applicants) are fully responsible for the experiments and the research workers involved, and that no more animals are used than indispensable, that no more pain, suffering and harms are caused than indispensable, that the staff is sufficiently skilled and that proper care for the animal subjects is provided (Section 9, para. 3, *German Animal Welfare Act*).
- Principal investigators are accountable for the use of experimental Refinements such as anesthesia, analgesia, humane endpoints and adequate killing methods (Sections 8 and 9, *German Animal Welfare Act; Appendix 1 in regards to No. 6.1.1, General Administrative Provision for the Execution of the Animal Welfare Act*).
- Care and housing, including the supervision of the animals and their veterinary care must be provided (Section 8, para. 3, No. 4, *German Animal Welfare Act*).

The amendments of the German Animal Welfare Act in 1998 did not change these obligations of animal researchers. The newest revision of the law in 2013, similarly, did not change in regards to the provisions to use experimental Refinements. Additional guidance on how to apply the law was provided in the General Administrative Provision for the Execution of the Animal Welfare Act of 9 February 2000 (Allgemeine Verwaltungsvorschrift zur Durchführung des Tierschutzgesetzes vom 9. Februar 2000) and built the basis of the application form to be used by animal researchers in 2010.

Animal research proposal

The following sections of the license proposals from 2010 were relevant to our assessment:

1. Section 1.6 describes the planned animal experiments including anesthesia (referring to Section 8, para. 2, sentence 3 in conjunction with Section 8a, para. 2 No. 3, *German Animal Welfare Act* of 1998). More specifically, the researcher has to elaborate on the kinds of interventions and treatments, their conduct and duration. Interventions or treatments that will be conducted under anesthesia must be specified including details about anesthetic regimens that will be employed. If painful procedures or treatments without anesthesia are to be conducted, this must be explained, and scientific justification must be given. Furthermore, if it is planned that there will be multiple severely painful interventions or treatments in an animal who is not anesthetized, this needs special scientific justification. The burden that the animals will have to endure (intensity and duration of pain and suffering) and harms inflicted upon them, including establishment of “humane endpoints” in case the proposed degree of severity will be exceeded, must be described in detail. Moreover, the planned actions need to be outlined that will be taken to reduce pain after anesthesia subsides (including agent, dosage, interval, duration).
2. In addition, a table needs to be completed, listing all procedures the animals will be subjected to and their expected burden, duration (in days), and intensity (severity classification for recovery experiments: mild, moderate, severe).
3. *The ethical justification for the research study must be given in section 1.7. The applicant must demonstrate that the expected pain, suffering or harm that is going to be inflicted upon the animals is ethically justifiable on the basis of the hoped-for benefits of the proposed research study.*
4. The intended whereabouts of the animals after the experiments are finalized has to be described in section 3.1 of the proposal.
5. If animals are to be killed, section 3.2 needs to describe which killing method is intended to be used. If injectables are to be used, the agent and dosage should be provided.

In 2010, it was not yet the norm to submit a health score sheet as part of the proposal. However, some competent authorities already requested these as they provide important information when assessing the degree of monitoring, care and treatment the animals would receive to ensure that pain and suffering is kept to an absolute minimum. To inflict pain, suffering or harm is only allowed to the degree that it is indispensable for reaching the experimental objective. The law underlines that it is not acceptable to inflict suffering to reduce the workload or save money (Section 9, para. 2, No. 3, *German Animal Welfare Act* as amended in 1998).

Task of the competent authority
The competent authorities are bodies put in place by the German government (as well as other member states of the European Union) to carry out the obligations arising from the law governing animal experimentation. Germany has 16 federal states and each state has, depending in its size and organizational structure, one or more competent authorities. These authorities are responsible for assessing animal experimentation proposals and, in the case of favorable project evaluations, authorize these. A newly submitted animal research application is reviewed by a veterinarian who is assisted by an animal experimentation committee according to Section 15 (German *Animal Welfare Act* as amended in 1986). This committee consists of scientists (2/3 of its members) and persons from animal welfare organizations (1/3 of its members) and ought to provide advice in regards to the legal requirements that need to be fulfilled in order to authorize animal experiments. In particular, this refers to the application of the 3Rs, indispensability of the project, and harm-benefit analysis (von Dehn and Nobiling, 2014). The competent authority's task is to scrutinize whether the proposal fulfills the legal prerequisites for experimenting on animals. After a proposal's first assessment by the competent authority and its committee according to Section 15 (*German Animal Welfare Act* as amended in 1986) the applicant oftentimes receives additional questions before a final decision (e.g., because information is missing or not clearly described).

Proposal selection criteria
Mice and rats (and more recently fish) are the most commonly used species in Germany (*Bundesministerium für Ernährung und Landwirtschaft, 2014, 2017*). The two rodent species are also among the most utilized animals in other countries of the European Union (European Commission, 2013). For this reason, our study was narrowed down to mice and rats. In total, 506 applications fit our criteria and, hence, were included in the analysis. The selected applications comprised 684 recovery surgical procedures (422 in mice and 262 in rats).

Research proposals that we included in this study met all of the following criteria:

1. Animal research proposals submitted to German competent authorities in 2010
2. Basic or applied research studies
3. Species: mice or rats
4. Procedures: surgical, recovery
5. Original proposals that were eventually granted a project license

Research proposals that we excluded from this study met one or more of the following criteria:

1. Applications for the generation of genetically altered mice
2. Government-required animal testing, e.g. toxicity testing of drugs, vaccines, pesticides
3. Species other than mice or rats
4. Procedures that did not involve recovery surgery
5. Proposals pre-assessed or rejected by members of the competent authorities

An electronic database was developed to compare and analyze the large volume of data collected. The database comprises the following information/fields:

1. Application number
2. Species
3. Type of surgical intervention
4. Name of specific surgery
5. Anesthetic regimens: A) Inhalational; B) Injectable; C) Combination of inhalational and injectable agents; D) Route of administration; E) Intubation: yes/no
6. Perioperative analgesic regimens: A) Names of agent(s); B) Time of analgesic administration: pre-operatively; at the end of surgery; together with anesthetic; no information; no analgesia
7. Structured pain assessments? yes/no
8. Severity classified: A) By investigator; B) According to guidance documents
9. Post-operative analgesia: A) yes; multimodal; no; no information; if needed according to the investigator; B) Type of analgesic; route of administration; dosage; application interval; duration; scientific explanation
10. Health Score Sheet in place? yes/no
11. Humane endpoints? yes/no
12. Killing: A) Under anesthesia, including type of anesthetic agent and dosage; B) Without prior anesthesia; C) Killing method; D) Scientific explanation

We examined the applications as originally submitted by the investigators so that we could assess the intended use of Refinement methods. This should demonstrate the investigators' understanding of legal requirements. Further, it should reflect their knowledge of potential Refinement methods, their understanding of how Refinement contributes to improved welfare of animals used in laboratories, and how Refinement is a prerequisite for obtaining rigorous data (Howard, 2013; Lloyd, Foden and Wolfensohn, 2008; Prescott and Lidster, 2017; Russell and Burch, 1959). All research applications in this review were eventually granted a license. Prior to the approval of a license, the competent authorities reviewed the proposals. In this process, it is expected that changes to the proposed protocols would be requested to comply with legislation. We did not review the follow-up questions that the competent authority members may have posed. To follow up on the outcomes of the assessments of the competent authorities and on their required modifications was not feasible as it would have entailed reviewing communication between the competent authority and applicant. Due to a shortage in personnel, it was impossible for the competent authorities to provide this additional anonymized paperwork. It also fell outside the scope of this study, which was to appraise the animal researchers' awareness of and know-how in regards to the application of Refinement methods.

2 Publications

2.1 Retrospective review of anesthetic and analgesic regimens used in animal research proposals

Citation: Herrmann, K. and Flecknell, P.A. (2019). Retrospective review of anesthetic and analgesic regimens used in animal research proposals. Alternatives to Animal Experimentation, 36(1), pp. 65-80. https://doi.org/10.14573/altex.1804011

Review Article

Retrospective Review of Anesthetic and Analgesic Regimens Used in Animal Research Proposals

Kathrin Herrmann [1,2] *and Paul Flecknell* [3]

[1]Freie Universität Berlin, Department of Veterinary Medicine, Institute of Pharmacology and Toxicology, Berlin, Germany; [2]current address: Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA; [3]University of Newcastle, The Medical School, Comparative Biology Centre, Newcastle upon Tyne, UK

Abstract

Pain has a profound effect on an animal's wellbeing. In Germany, researchers using animals have been legally required to reduce any possible pain, suffering, distress or lasting harm to an absolute minimum since 1972. To evaluate how these provisions have been implemented in practice, an assessment of refinements to experimental techniques was conducted by retrospectively reviewing 684 surgical interventions described in 506 animal research applications that were sent to the German competent authorities for approval in 2010. This paper focuses on the efficacy of proposed anesthesia and peri- and postoperative analgesia. Postoperative analgesia was not proposed for 30% of surgeries. Following 10% of procedures, animals were to be given pain relieving medication if the investigators decided this was necessary; however, structured assessments to detect pain were absent. Consequences of unalleviated pain and omission of pain assessment techniques are discussed, and some recommendations to improve anesthesia and analgesia are given. The findings of this review highlight the need for improvement, both to fulfil legal requirements and to improve animal welfare. To monitor compliance with animal welfare regulations and ensure good veterinary and scientific practices, education and training need to be intensified. Adherence to the items listed in the PREPARE and ARRIVE guidelines and the Gold Standard Publication Checklist (GSPC) should become legally binding.

1 Introduction

Refinement, the last R of the Three Rs principles (replacement, reduction, refinement), was first described by Russell and Burch 60 years ago (Russell and Burch, 1959), and pertains to "*[m]ethods that minimize any pain, suffering, distress or lasting harm that may be experienced by the animals, and improve animal welfare. Refinement applies to all aspects of animal use, from the housing and husbandry to the scientific procedures performed upon them*" (Graham and Prescott, 2015). The revised Directive, 2010/63/EU, requires European Union (EU) Member States to fully implement the Three Rs principles in their national laws (EU, 2010). Most states amended their national legislation in 2013. Well before that, following the revision of the German Animal Welfare Act in 1972, research workers using animals in Germany were required to reduce any possible pain suffering, distress, or lasting harm caused to animals to an absolute minimum (Germany, 1972). Nevertheless, due to limitations in the transparency of the authorization processes and reporting of animal research, it has been difficult to assess the extent to which experimental procedures were refined in practice.

Structured literature reviews of research that involved the use of laboratory animals have provided some insights. Animal research involving surgical procedures carried out on pigs, sheep, dogs, and non-human primates (Coulter et al., 2009), rabbits (Coulter et al., 2011) and rodents (Coulter et al., 2009; Richardson and Flecknell, 2005), published in peer-reviewed journals, has been analyzed with regard to analgesic and anesthetic administration. Stokes and colleagues (2009) focused on studies conducted during two time periods (2000-2001 and 2005-2006) to assess changes in the administration of analgesics and anesthetics to laboratory mice and rats undergoing surgical procedures. The study showed a trend towards improvement such as safer anesthetic regimens used in the later period examined, but this and an earlier review assessing analgesic use in rodents (Richardson and Flecknell, 2005) found that there was still significant scope for refinement, especially with respect to perioperative care. Richardson and Flecknell (2005) not only screened

Received April 1, 2018; Accepted September 13, 2018;
Epub September 14, 2018; © The Authors, 2018.

ALTEX 36(1), 65-80. doi:10.14573/altex.1804011

Correspondence: Kathrin Herrmann, Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT)
Baltimore, MD 21205, USA
(kherrma1@jhu.edu)

the publications for postoperative analgesia but also e-mailed the authors or the respective Animal Ethics Committees a survey to find out when analgesia was administered independent of whether it was reported. The response rate was low (28 out of 101 papers) and only 18% of correspondents had given analgesics but not reported it. A systematic review of anesthesia, analgesia, and euthanasia methods used in anesthesiology, respiratory, and critical care research published in the ten highest ranked journals showed insufficient reporting of treatment regimens in experimental studies using small laboratory mammals; for example, the administration of analgesics was only reported in 19% of all painful interventions (Uhlig et al., 2015).

An assessment of papers published before 2011 and from 2014 to 2015 further confirmed that reporting of experimental refinement methods is still poor (Carbone and Austin, 2016). Whilst guidance documents specially written for reporting of animal studies, e.g., the Gold Standard Publication Checklist (GSPC) (Hooijmans et al., 2010) and the ARRIVE guidelines (Kilkenny et al., 2010), have been available since 2010, scientific publications still cannot be relied upon to present a detailed description of analgesia and anesthesia protocols. Most recently, an assessment of anesthetic and analgesic regimens in publications on studies involving non-human primates once again confirmed the lack of reporting of critical detail (Bertrand et al., 2018). Although many journals have adopted guidelines such as "ARRIVE", publications continue to omit key details of study design and conduct (e.g., Leung et al., 2018). This may be due to a lack of appreciation by researchers of the need to include these data in their publications. It therefore remains unclear to what extent the experimenters did not report their refinement practices although, for example, proper pain management may have been provided.

Besides the lack of information regarding refinement in the available publications, there is a general lack of transparency within animal research, as in all European Union Member States access to animal research proposals is generally restricted to members of the regulatory authorities. Thus, the degree to which refinement methods are applied in practice and the awareness of refinement options by research workers remains uncertain.

German animal research proposals must include detailed descriptions on how all procedures and interventions are to be conducted, including precise descriptions of surgeries and the application of refinements. The most humane and refined methods have to be used while still ensuring valid data collection; this is facilitated by a robust and effective harm-benefit analysis. For example, if less humane anesthetics are to be used or analgesia is withheld, this must be scientifically justified, and the hoped-for benefits of the research must outweigh also the additional harm.

To evaluate the use of experimental refinements in German laboratories, we retrospectively reviewed original animal research applications submitted to the German regulatory authorities for approval in 2010. The efficacy of the proposed anesthetic and peri- and postoperative analgesic regimens, the health monitoring, use of humane endpoints, killing methods, and the researchers' categorization of the severity of planned procedures were investigated. The goal was to assess the application of refinements before Directive 2010/63/EU came into effect and to make recommendations about where further improvements could be made. To our knowledge, this is the first assessment of its kind. In this paper, we focus on proposed anesthetic and peri- and postoperative analgesic regimens and discuss deficiencies and necessary improvements.

2 Materials and methods

In Germany, reviewing and licensing animal research proposals falls under the authority of each of the federal states. The 16 federal states have their own "competent authorities" for this task, generally comprised of state veterinarians. After negotiations with the highest authorities of the federal states, 14 of the 16 agreed to provide us access to original proposals for basic and applied research that had been granted a license. For the scope of our study, we selected all applications submitted to the participating competent authorities in 2010 that included recovery surgical procedures in mice and rats to assess appropriateness of anesthesia, perioperative analgesia, and postoperative care. Mice and rats are the two most commonly used species in Germany; over 2,000,000 mice and around 400,000 rats per year are used for scientific purposes in Germany (BMEL, 2014, 2017). Elsewhere in Europe, these rodents are also among the most frequently used species (EC, 2013; Home Office, 2017). A total of 506 applications met these criteria and were included in the analysis. These included 684 recovery surgical procedures (422 in mice and 262 in rats). The study was conducted anonymously (i.e., the individual research groups were not identified in the analysis).

Research proposals were included in this study if they met the following criteria:

1. Animal research proposal submitted to a German competent authority in 2010
2. Basic or applied research study
3. Species: mouse or rat
4. Procedure: surgical, recovery
5. Original proposal that was granted a project license

Research proposals were excluded from this study if they met the following criteria:

1. Application for the generation of genetically altered mice
2. Government-required animal testing, e.g., toxicity testing of drugs, vaccines, pesticides
3. Species other than mouse and rat
4. Procedure that did not involve recovery surgery
5. Proposal pre-assessed or rejected by members of the competent authority

We analyzed the experimental protocols described in the original applications. However, before approval of a final license, it is possible, and in some cases very likely, that the competent authorities requested changes to the proposed protocols. To follow up on the results of the competent authorities' assessments and on the amendments made in the process was not feasible. Also, the scope of our study was to evaluate the researchers' awareness of and expertise in the application of refinement methods, not that of the competent authorities.

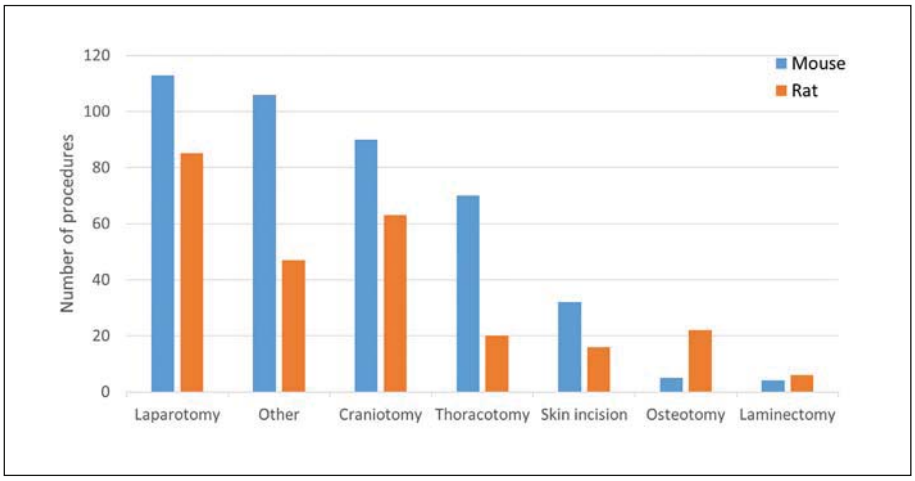


Fig. 1: Types and numbers of surgical procedures performed on mice and rats
A total of 684 surgical procedures were performed, 422 on mice and 262 on rats. "Other" refers to a general category of all surgeries that did not involve the opening of a body cavity, e.g., compression or ligation of nerves, or surgery to cause a middle cerebral artery occlusion.

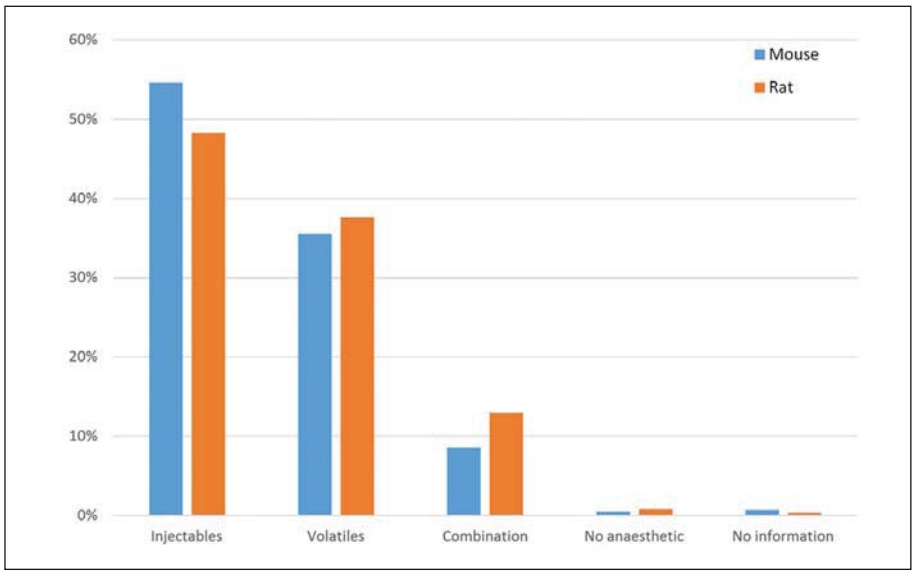


Fig. 2: Anesthetic regimens used
A total of 684 anesthetic regimens were performed, 422 on mice and 262 on rats.

3 Results

Some applications lacked required information. For example, in a few cases the type, the dose, or the route of administration of an anesthetic agent was missing; the justification as to why analgesia was withheld was mostly absent and there was generally no discussion on why certain anesthetic regimes were selected, even when they did not meet the standards of good veterinary practice. Project licenses were generally granted for up to three years, with the option to extend up to five years.

3.1 Types of surgical interventions and frequency of endotracheal intubations

The 506 assessed applications comprised 684 surgical procedures. Some animals underwent more than one surgery or in some cases several experimental groups underwent different surgical procedures. Out of the 684 surgical procedures, 422 were performed on mice and 262 on rats.

The most frequently performed surgical procedures on mice and rats were laparotomies. For rats, these were followed by craniotomies. For mice, the second most common procedure was surgery that did not involve the opening of a body cavity (Fig. 1); examples include neuropathic pain models such as compression or ligation of nerves (e.g., ischiatic nerve), implantation of catheters in veins and arteries (e.g., femoral artery or vein, jugular vein), a hindlimb ischemia model (ligation of the femoral artery), and a focal cerebral ischemia model (middle cerebral artery occlusion).

Endotracheal intubation was planned for 15% of surgeries in mice, the vast majority of these (82%) for thoracotomies, and for 8% of surgeries on rats, over half of these being for thoracotomies.

3.2 Anesthetic regimens and agents used

Injectable anesthetics were the most commonly used anesthetics in procedures on both mice (55%) and rats (48%), followed by inhalant agents (mice, 36%; rats, 38%) (Fig. 2). A combination

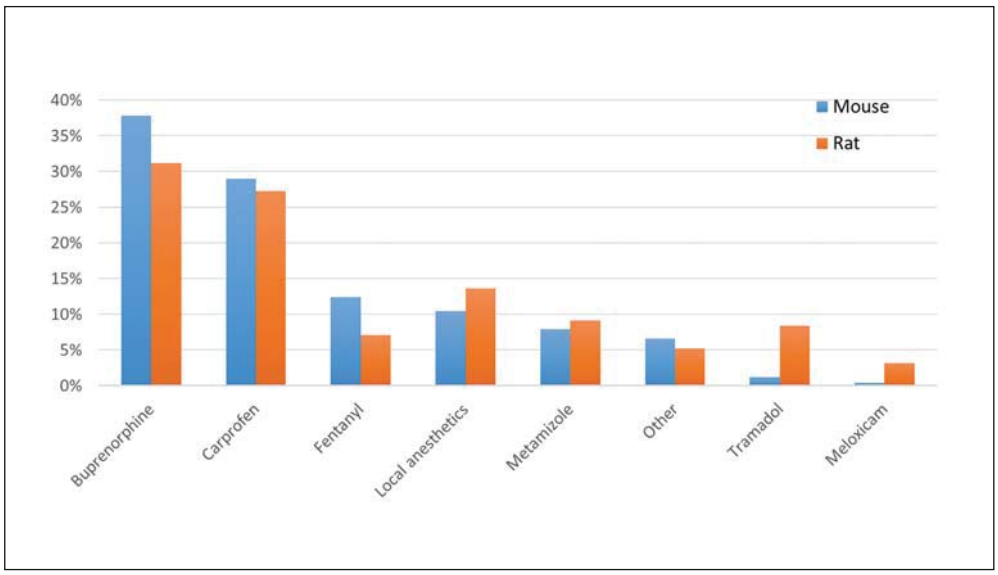


Fig. 3: Percentage of analgesia agents and local anesthetics given perioperatively to mice and rats
Lidocaine, xylocaine, ropivacaine, bupivacaine, and amethocaine were used as local anesthetics. "Other" agents used included paracetamol (acetaminophen), flunixin meglumine, and butorphanol.

of injectable and inhalational agents was used in about 10% of surgeries (Fig. 2). The authors did not provide information on anesthesia in 1% of surgeries. For another 1% of the surgical interventions it was not planned to use an anesthetic agent. Instead, the investigators intended to use hypothermia (cooling the animals to 1-4°C by laying them on ice for about six minutes) to immobilize neonatal mice and rats while carrying out intracranial injections of virions or neuroblasts.

Injectable agents used for anesthesia
Ketamine and xylazine were the most frequently used injectables in mice (79%) and in rats (70%). A completely reversible anesthetic regimen, a mixture of fentanyl, medetomidine, and midazolam, was used in 11% of mice and in 5% of rats. Choral hydrate was frequently used in rats (6%) but was rarely administered to mice (< 1% of procedures). Ketamine and medetomidine were also frequently used in rats (6%) rather than in mice (2%). Pentobarbital use ranked between 3% in mice and 5% in rats. Other single agents used in mice were medetomidine (1%) and ketamine (< 1%). Acepromazine was combined with xylazine and ketamine for rats (5%). In mice, acepromazine was combined with ketamine but this was used rarely (< 1%). The mixture of fentanyl, fluanisone, and midazolam was also used rarely for mouse and rat surgeries (2%). In a few cases it was stated that reversible anesthetic agents would be antagonized at the end of the procedure.

Inhalational agents
Isoflurane was by far the most widely used inhalation agent (74% and 70% of procedures performed in mice and rats, respectively), followed by the combination of isoflurane and nitrous oxide in both species (15%). For surgeries on mice, nitrous oxide was also used in combination with halothane (9%) as well as halothane alone (1%), whereas for rat surgical interventions no halothane use was reported, but sevoflurane (8%) and methoxyflurane (7%) were used.

Combination of inhalational and injectable agents producing anesthesia
The combination of inhalational agents with injectable anesthetics or sedatives was used in 9% of surgeries on mice and 13% of surgeries on rats. Here the most frequently administered injectable agents – ketamine and xylazine – were combined with isoflurane: 39% for mice; 56% for rats. The next most commonly used combination of anesthetics for mice was isoflurane and ketamine/midazolam and for rats isoflurane/nitrous oxide with ketamine/xylazine. In the majority (61%) of cases anesthetic induction was performed with the injectable anesthetic/sedative, followed by the use of inhalational agents. In some cases, ether (3% of cases) or carbon dioxide (6% of cases) was used to initiate anesthesia in rats. Carbon dioxide was however frequently used to kill animals (Herrmann and Flecknell, 2018a).

3.3 Perioperative analgesia
The most frequently used perioperative analgesic agent for both mice and rats was buprenorphine (mice: 38%; rats: 31%), followed by carprofen (mice: 29%; rats 27%), and fentanyl (mice: 12%; rats: 7%) (Fig. 3). Local anesthetics such as lidocaine, xylocaine, ropivacaine, bupivacaine, and amethocaine were used in 10-13% of procedures. For instance, 22% of the craniotomies and 10% of the thoracotomies involved the application of local anesthetics at the surgical site. Metamizole was used in 8-9% of cases, followed by other analgesics such as paracetamol, flunixin meglumine, and butorphanol. Tramadol was used rarely in mice, but was administered after 8% of surgeries on rats. Meloxicam was a seldom administered agent in both species (Fig. 3).

When inhalational anesthetics with no or minimal analgesic properties were used, it was assessed whether and when they were combined with analgesics. In the case of isoflurane, 25% of mice and 28% of rats received no pain relief, and 13% of mice and 18% of rats received pain relief at the end of surgery, while the others received pain relief pre- or intraoperatively (Fig. 4).

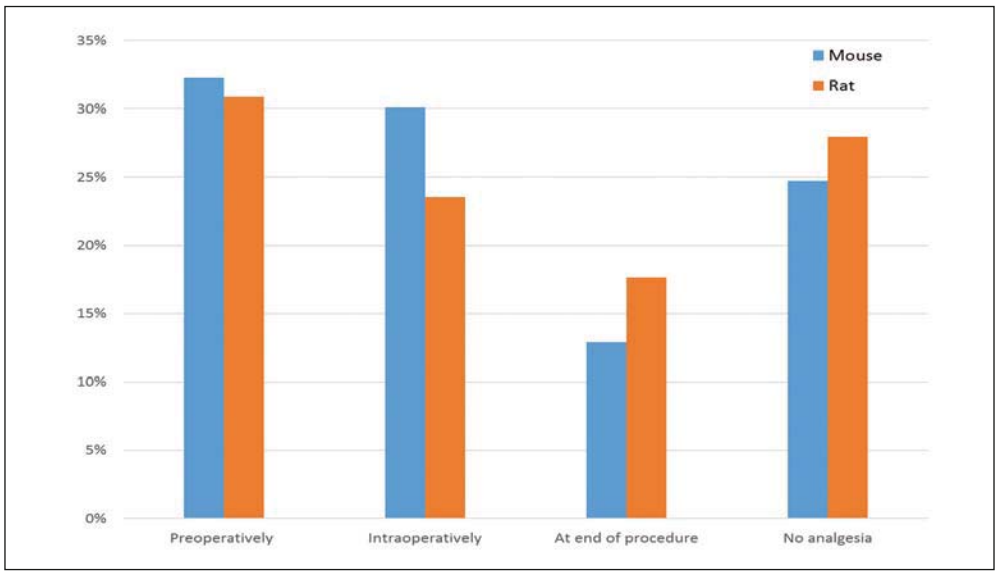


Fig. 4: Time of analgesic administration when isoflurane was used
Of all inhalational anesthetic regimens, isoflurane was used in 74% and 70% of surgical interventions in mice and rats, respectively.

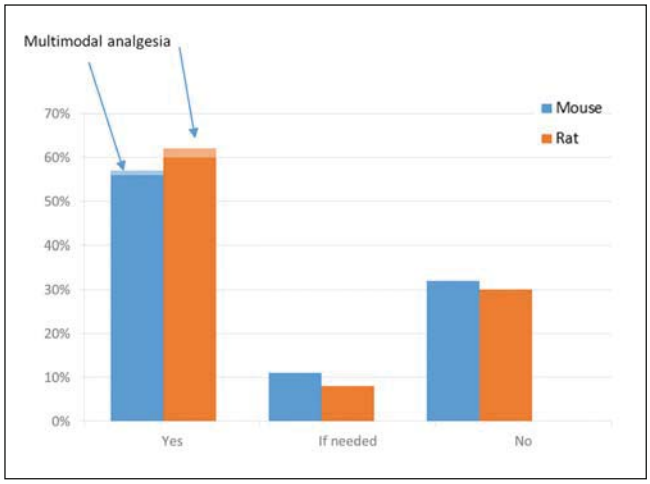


Fig. 5: Decision on administration of analgesic treatment following surgical interventions
(422 on mice and 262 on rats): yes, no, and "if needed" according to the judgement of the researcher.

For injectable agents with no specific analgesic activity such as pentobarbital or chloral hydrate we reviewed whether these were combined with analgesics: With pentobarbital, analgesia was not provided in 55% of surgical procedures on mice and in 75% of surgical procedures on rats. Chloral hydrate, an injectable anesthetic that was most frequently used for craniotomies in rats, was most often combined with tramadol given intraoperatively and, less often, with a local anesthetic.

3.4 Postoperative analgesia

Postoperative analgesia was administered following 57% of the surgical interventions on mice, less than 1% being multimodal, and 62% of procedures on rats. In only 1% and less than 2% of cases, respectively, were these multimodal analgesic regimens (Fig. 5).

The surgical procedures for which multimodal analgesia was provided postoperatively are listed in Table 1 (for mice) and 2 (for rats). In one of the two mouse procedures, the multimodal approach was optional, depending on pain being detected. The first three postsurgical regimens in Table 2 are identical and were used after various surgical procedures all described in the same research application. One of the aortic banding surgeries, a commonly used experimental model for pressure overload-induced cardiac hypertrophy and heart failure, used the multimodal approach only in case the animals did not drink enough, because metamizole was supplied via the drinking water, and/or the animals showed signs of pain. The applications involving aortic banding (Tab. 2) represent two of very few that mentioned they would monitor water intake. One application also stated the amount that the rats were expected to drink from 8 weeks old (8-12 ml/kg body weight).

No postoperative analgesia was intended following 32% of surgeries on mice and 30% of surgeries on rats (Fig. 5). Reasons for withholding analgesia were given for 24% of these surgeries on mice and 33% of these surgeries on rats. The most frequent explanations were that it was not considered necessary as pain was anticipated to be minor, and that the analgesic agent could influence study outcomes. In 11% of surgical procedures on mice and in 8% of surgical procedures on rats the investigators declared that they would administer analgesia only "if needed" (Fig. 5). Pain management would be considered "when signs of pain and/or suffering were observed" or "in case of signs of moderate pain". Some stated that, according to their experience, analgesia was not necessary for the planned surgical interventions and that the surgery was "tolerated well by the animals". Overall, these applications contained no description of how a structured pain assessment would be conducted. The general health score sheets 15% of surgical procedures had included focused on various health indicators, including humane endpoints, most of which are not specific for pain (Herrmann and Flecknell, 2018a,b).

Similar to intraoperative pain management for mice and rats, buprenorphine (mice: 33%; rats: 34%) and carprofen (both: 33%)

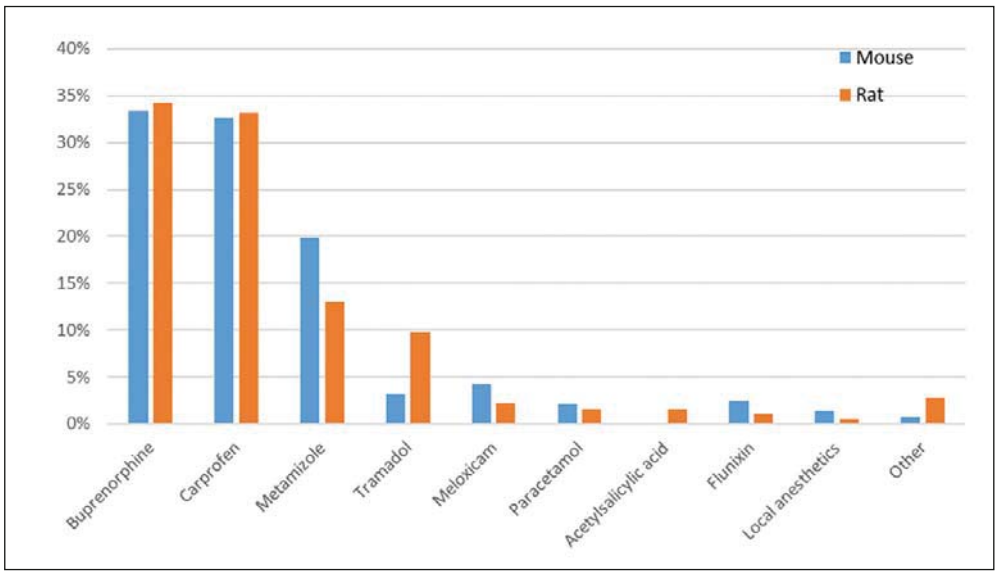


Fig. 6: Postoperative analgesic agents used
No analgesia was provided after 32% of mouse surgeries and after 30% of rat surgeries. Proposed local anesthetics were lidocaine, xylocaine and bupivacaine. Other analgesics included piritramide, fentanyl, ketamine and ketoprofen.

were most commonly utilized to manage postoperative pain, followed by metamizole (mice: 20%; rats: 13%) (Fig. 6). Tramadol was used postoperatively in 10% of surgeries in rats and in 3% of surgeries in mice. Meloxicam was more popular after interventions in mice (4%) than in rats (2%). Other rarely used agents were paracetamol, acetylsalicylic acid, flunixin meglumine, and local anesthetics (Fig. 6).

Information regarding duration and frequency of postoperative analgesic application was incomplete. When indicated, buprenorphine was often given subcutaneously (s.c.) between once and three times a day for one to three days and longer, but some administered buprenorphine intraperitoneally (i.p.) or intramuscularly (i.m.). Carprofen was given mostly s.c., but at times i.p., i.m., orally, or via the drinking water. When not given via the drinking water, it was administered once or twice daily for one to three days, or occasionally longer. Metamizole, paracetamol, tramadol, and acetylsalicylic acid were usually given via the drinking water. Duration of administration of metamizole varied between one to five days, in some cases up to seven days and longer. Paracetamol was given mostly for two to three days. For tramadol, the duration was rarely mentioned and for acetylsalicylic acid not at all. The route of administration of local anesthetics was mostly not specified.

Analgesic doses given were usually within the standard published ranges with the occasional low, e.g., 2.5 mg/kg carprofen in rats, or high, e.g., 30 mg/kg of carprofen in mice. Meloxicam was dosed mostly at 0.2 mg/kg. Dosages for metamizole supplied via the drinking water ranged between 1.3 mg/ml and 2.5 mg/ml. Other doses noted were 110 mg/kg, 200 mg/kg, and 300 mg/kg. Paracetamol was provided at 1-3 mg/ml drinking water. Tramadol drinking water doses were 30 to 50 mg/kg/8h or 25 mg/l up to 200 mg/l drinking water. It was rarely reported whether the drinking water intake would be measured, and how much an animal, based on body weight, was expected to drink to assure sufficient drug levels. A few investigators stated that they would administer the analgesic s.c. if animals showed pain despite having the agent in their drinking water. In one proposal, administration of tramadol via the drinking water was started preoperatively. A small number of investigators stated that they would switch to another analgesic agent or give an additional analgesic if pain was not controlled effectively. Some asserted that they would provide analgesia until no more pain was observed. However, information in regard to performing a structured pain assessment was absent.

3.5 Health and pain monitoring

Health score sheets were available for 108 out of 684 surgical procedures (15%); their quality varied. The vast majority monitored general indicators of deteriorating health not specific to pain – such as weight loss, alterations in body temperature, and tumor size. When declaring that the animal's behavior and overall health condition would be monitored, it was mostly unclear to what extent this would be done and what would be done to alleviate clinical signs (analgesia, special bedding or other measures of intensive care). The planned frequency of monitoring of the animals was mentioned for 33% of all surgical interventions, and in the majority of cases these checks were to be done once a day. More results on health monitoring, severity of procedures, and assessment outcomes with regard to humane study endpoints and killing methods are published elsewhere (Herrmann and Flecknell, 2018a,b).

4 Discussion

We found shortcomings in all areas we assessed. First and foremost, pain was often not alleviated or was unlikely to be alleviated effectively. Following 30% of surgical interventions, no analgesia was planned, and in another 10% of procedures it would be alleviated only if pain was observed but monitoring generally seemed to be infrequent or lacking. Reasons for withholding analgesia, if included, may indicate not only a general lack of monitoring, a lack of pain assessment, and a lack of pain recognition skills; they also imply that concerns about the neg-

Tab. 1: Post-procedural multimodal analgesic approaches following surgery on mice
Postoperative analgesia was administered following 57% of the surgical interventions on mice, less than 1% being multimodal.

Type of surgical intervention	Intraoperative regimen	Multimodal regimen
Sternotomy; cryoinjury-induced myocardial infarction	Induction: isoflurane 1.5-3 Vol%; maintenance: isoflurane 0.8-1 Vol%; nitrous oxide 60%; O2 40%; intubation; intercostal space infiltration with ropivacaine; buprenorphine, 0.1 mg/kg i.p.; carprofen, 5 mg/kg s.c.	Buprenorphine, 0.1 mg/kg i.p., 3x/d, duration: 5 d; carprofen, 5 mg/kg s.c., 1x/d, duration: 10 d
Craniotomy; chronic cranial window preparation (bone flap removal and replacement by glass cover)	Ketamine, 130 mg/kg and xylazine, 10 mg/kg i.p.; carprofen, 5 mg/kg s.c., buprenorphine, 0.05 mg/kg i.p.	Carprofen, 5 mg/kg s.c.; duration: 3 d; if pain detected: buprenorphine, 0.05 mg/kg i.p.

Tab. 2: Post-procedural multimodal analgesic approaches following surgical interventions on rats
Postoperative analgesia was administered following 62% of procedures on rats, 2% being multimodal.

Type of surgical intervention	Intraoperative regimen	Multimodal regimen
Laparotomy; clamping of Ligamentum hepatoduodenale (Pringle maneuver); ischemia for 30 minutes	Buprenorphine, 0.05-0.1 mg/kg s.c.; isoflurane: induction, 4-5 Vol%; maintenance, 1.5-3 Vol%	Carprofen, 5-10 mg/kg s.c. 1x/d; buprenorphine* s.c. 2x/d, duration: 3 d
Laparotomy; 2/3-hepatoectomy (in some groups combined with 30 minute-Pringle maneuver)	Buprenorphine, 0.05-0.1 mg/kg s.c.; isoflurane: induction, 4-5 Vol%; maintenance, 1.5-3 Vol%	Carprofen, 5-10 mg/kg s.c. 1x/d; buprenorphine* s.c. 2x/d, duration: 3 d
Laparotomy; placing of gastrointestinal tube	Ketamine, 40-50 mg/kg and xylazine, 1-2 mg/kg i.p.	Carprofen, 5-10 mg/kg s.c. 1x/d; buprenorphine* s.c. 2x/d, duration: 3 d
Partial sternotomy; aortic banding	Ketamine, 80-100 mg/kg and xylazine, 1.5-5 mg/kg i.p.; injection of carprofen* at the beginning (route not further specified); intubation; injection of metamizole* at end of procedure (route not further specified)	Metamizole* via drinking water for 3 d; carprofen* s.c. 1 d (longer if needed)
Partial sternotomy; aortic banding	Ketamine, 80-100 mg/kg and xylazine, 1.5-5 mg/kg i.p.; intubation; metamizole* s.c. at end of procedure	Metamizole, 0.05 g in 50 ml drinking water mixed with 0.25 ml 20% glucose solution for 3 d; if needed additionally carprofen* s.c. 1x/d
Placement of titanium implant on two critical size defects (5 mm diameter) per scull	Ketamine, 90 mg/kg and xylazine, 10 mg/kg i.p.	Carprofen, 5 mg/kg 2x/d; buprenorphine, 0.05-0.1 mg/kg 2x/d, both s.c.; duration: 1 d, max. 2 d

*no dose given

ative effects of analgesics on study outcomes were greater than the concerns about untreated pain.

Considering the severity of planned surgical interventions, it was unexpected that use of multimodal analgesia was almost non-existent. Furthermore, preventive analgesia was often not provided as, for example, anesthetics that have no analgesic properties (e.g., isoflurane) were to be used; this was the case in a quarter of these procedures, and in another 13-18%, analgesia was planned to be given too late, and thus, the animals would recover consciousness without pain relief. Moreover, the use of local anesthetics for severe surgeries such as thoracotomies was very low (10% of thoracotomies).

In the following, we discuss the problems we identified in detail and provide some information as to how these could be improved or resolved. We are confident that implementation of these refinements would both significantly improve animal welfare and improve the quality of data obtained.

4.1 Unalleviated pain

As we found that pain management during and after surgery was often absent or inadequate, we commence the discussion with general considerations on the effects of unalleviated pain. Researchers often appear to be concerned that (side)effects produced by analgesics (and anesthetics) could potentially influence or even confound study outcomes and data (as also seen in our study sample). However, there is an abundance of information on the pharmacodynamics, pharmacokinetics, unwanted effects, toxicity, and interactions of analgesics and anesthetics with other drugs (e.g., Flecknell, 2016; Gaynor and Muir, 2014; Grimm et al., 2015; Muir and Hubbell, 2014; Informationssystem CliniPharm CliniTox[1]) to

[1] CliniPharm Wirkstoffdaten. http://www.vetpharm.uzh.ch/perldocs/index_i.htm

guide investigators when designing their studies. In addition, there is an increasing number of recommendations for the refinement of certain models and procedures, for example, for models of ischemic stroke (Percie du Sert et al., 2017), for rheumatoid arthritis (Hawkins et al., 2015), for sepsis and septic shock (Lilley et al., 2015), for experimental autoimmune encephalomyelitis (EAE) (Wolfensohn et al., 2013a), and for procedures involving seizures, convulsions, and epilepsy (Lidster et al., 2016; Wolfensohn et al., 2013b). These guidelines were composed by experts based on their knowledge of the particular models and the value of potential refinements. Furthermore, specialists in veterinary anesthesia and analgesia in a research setting can give additional, project-specific advice and guidance.

Unalleviated pain has profound effects on the animal, leading to emotional, behavioral, and physiological changes (see Jirkof, 2017; Peterson et al., 2017). The immune system is suppressed due to pain stimulating glucocorticoid and catecholamine release. Responses to pain vary between individuals, and consequently are hard to predict, which potentially makes pain a poorly controlled study variable. In comparison, the influence of anesthetic and analgesic agents on the data is more foreseeable than the effects of unalleviated pain, distress, and suffering on *in vivo* research. In addition, use of perioperative analgesics – if based on the need of the individual animal elicited by properly and frequently conducted structured pain assessments – will most likely have relatively minor effects on the animal model. Confounding effects are more likely to be introduced by analgesics when the agents are given over an extended time period or in unnecessarily high dosages. The likely effects are often overestimated and based on studies in which analgesics were administered at elevated doses and for time periods longer than needed in most experimental settings (e.g., Hall et al., 1996). Furthermore, models designed to be of relevance to clinical conditions in humans should be treated as in the human setting, where pain would generally be alleviated whenever possible. A plan for adequate and controlled pain management should be developed as part of the study protocol. To sufficiently manage pain, the choice of anesthetics and analgesics, their dose, administration route, frequency and duration of treatment, and a pain assessment plan should be tailored to the individual research project, treatment group and, if necessary, to an individual animal's needs.

4.2 Need for refinement of anesthetic protocols

Endotracheal intubation

Intubation was performed in 8% and 15% of surgeries (rats and mice, respectively), and the majority of these were thoracotomies. When opening the thoracic cavity, endotracheal intubation is generally indicated as the animals require mechanical ventilation to maintain respiratory function. In addition, there are certain disease models where intubation and mechanical ventilation is of considerable benefit. For example, the lesion size in experimental models of stroke is influenced by cerebral blood flow and the degree of cerebral vasodilation. These variables are markedly affected by changes in arterial carbon dioxide, which can be maintained at a consistent level by mechanical ventilation (Liu et al., 2009). Tracheal intubation might appear challenging in small rodents but it is relatively easy to perform with the correct apparatus (Flecknell, 2016).

Hypothermia

For experiments where neonatal mice and rats were to be injected intracranially, the researchers planned to use hypothermia to immobilize them. The investigators would then inject the animals' brains with virions or neuroblasts. In neonatal rodents, the nociceptive pathways are thought to be not yet fully developed (Fitzgerald, 2005). However, recovery from hypothermia is possibly associated with pain as it has been shown to lead to c-Fos activation (Rhodes, 2009). Since there are anesthetic agents available for neonatal mice and rats such as methoxyflurane (Danneman and Mandrell, 1997) or fentanyl/fluanisone (Clowry and Flecknell, 2000), the use of hypothermia should be abandoned. It is our understanding that the German competent authorities generally do not allow this controversial practice (*Projektgruppe der Genehmigungsbehörden*, personal communications, 2016).

Pentobarbital

A number of surgical interventions were conducted under the hypnotic agent pentobarbital, given intraperitoneally. Doses ranged from 36 mg/kg body weight up to 60 mg/kg in mice and up to 70 mg/kg in rats whereas recommended doses are between 40-50 mg/kg (Field et al., 1993). Pentobarbital has a narrow safety margin and is best used to provide a light plane of anesthesia rather than for surgical interventions, for which doses close to the lethal dose may be required (Flecknell, 2016). The high pH of pentobarbital (11) causes pain when injected intraperitoneally. Pentobarbital also results in prolonged sleep time in rats (120-240 min) and mice (120-180 min) (Flecknell, 2016), and during this period, animals have depressed cardiovascular and respiratory systems and are susceptible to hypothermia.

Chloral hydrate

Choral hydrate was rarely used in mice, but was the second most frequently used agent in rats. However, it can cause adynamic ileus in rats (Fleischman et al., 1977) and it often produces only a light plane of anesthesia. Most investigators using choral hydrate planned to give perioperative analgesia. Nonetheless, due to the danger of causing adynamic ileus, chloral hydrate should be replaced by more effective and safer agents (Baxter et al., 2009).

Use of ketamine or medetomidine as the sole agent

There were few cases for which the use of ketamine or medetomidine as single agents was planned. Used alone, these agents cause sedation and mild to moderate analgesia, but they must be combined with other agents to produce surgical anesthesia. The combination of ketamine with acepromazine or with midazolam does not usually provide an anesthetic depth necessary for surgery in either species (Flecknell, 2016); hence, for surgical interventions, additional agents are needed.

Use of inappropriate inhalational agents for induction of anesthesia

For rats, ether was used in 3% and carbon dioxide in 6% of protocols to initiate anesthesia. Induction with both of these agents

is distressing and aversive due to irritation of the mucus membranes. Coughing, salivation, and at times laryngospasm occur from inhaling ether (Flecknell, 2016). Carbon dioxide is strongly aversive, and causes distress and fear in rodents (e.g., Chisholm et al., 2013; Kirkden et al., 2008; Leach et al., 2002a,b, 2004; Makowska and Weary, 2009; Marquardt, 2013; Niel and Weary, 2007; Niel et al., 2008; Wong et al., 2013; Ziemann et al., 2009) and thus, its use as a killing agent has been controversial for many years. Rats and mice are unwilling to tolerate exposure to CO_2 long enough to lose consciousness (Leach et al., 2002b, 2004; Niel and Weary, 2007; Niel et al., 2008). Food incentives to stay in the gas chamber (Leach et al., 2004; Niel and Weary, 2007) and withholding food for up to 24 hours to increase their motivation to stay in the chamber longer as well as varying the gas flow rates did not change the time they were willing to tolerate carbon dioxide (Niel et al., 2008). When comparing carbon dioxide to other anesthetic gases, it was found to be the most aversive gas for both mice and rats, with the least aversive being enflurane and halothane for mice and halothane for rats (Leach et al., 2002a). Other studies compared rats' aversion to carbon dioxide with isoflurane or halothane exposure and concluded that the fluorinated hydrocarbons were less aversive than carbon dioxide (Kirkden et al., 2008; Makowska and Weary, 2009; Wong et al., 2013; Young, 2006).

Recommendations

Dosing and route of administration

Drug doses vary between species, strains, and individuals and depend on age, sex, and health condition of the individual animal (Flecknell, 2016). Therefore, we only discuss chosen dosages if they were outside of the standard range. Injectable anesthetics were the most widely used agents in both mice and rats. They can provide safe and effective anesthesia, but this is most easily achieved when given intravenously, so that the dosage can be adjusted to provide the desired effect in that particular animal. However, this is generally not feasible for small rodents, who thus receive most injectable anesthetic agents intraperitoneally (i.p.) and less often subcutaneously (s.c.) or intramuscularly (i.m.). Several studies have shown that intraperitoneal applications have a high failure rate (6-20% for experienced personnel) with some of the material being injected, e.g., into the fatty tissues, viscera, or caecum (Ballard, 2009; Coria-Avila et al., 2007; Gaines Das and North, 2007; Lewis et al., 1966). Intramuscular injections can be painful, and in the case of ketamine, for example, they can produce necrosis (Smiler et al., 1990). The subcutaneous route seems the least invasive, however, data comparing the accuracy of dosing with i.p. administration is not available. Since injectable anesthetics are administered as a single dose, overdosing or underdosing can easily occur. To avoid complications, agents with a wide safety margin should be used as well as agents whose effects can be reversed using specific antagonists.

Reversible anesthetic agents

Due to their small body size, rodents are prone to hypothermia, and respiratory and cardiovascular depression (Fleischmann et al., 2016), and mortality increases with prolonged postsurgical sleep time. This risk can be reduced by using anesthetic agents that can be reversed such as medetomidine and xylazine with atipamezole; midazolam with flumazenil and fentanyl with butorphanol, buprenorphine or naloxone. Some researchers who used reversible anesthetic regimens planned to antagonize these at the end of the procedure. This should be done routinely to avoid long sleep times and ensure a faster recovery.

Most investigators in our study sample used ketamine and xylazine, a regimen associated with prolonged recovery periods (Albrecht et al., 2014). Ketamine/xylazine and ketamine/medetomidine could be partially antagonized with atipamezole; this should be considered to reduce recovery time.

Pre-medication and balanced anesthesia

Isoflurane was the most frequent inhalational anesthetic used. There seems to be a general trend towards using inhalant anesthetics, probably because induction is rapid, anesthesia levels can be adjusted to the individual animal's needs, and recovery from anesthesia is fast. Isoflurane undergoes almost no biotransformation; however, it is mildly pungent (Eger, 1981). Most commonly used inhalational agents are to some degree pungent to rodents, and studies have shown increased aversion and distress after repeated exposure to, e.g., isoflurane and halothane (Makowska and Weary, 2009). Hence, it is advisable to use pre-medication such as fentanyl/fluanisone or midazolam/fentanyl for induction before maintenance with an inhalational agent. Other commonly used pre-medication agents are medetomidine, dexmedetomidine, or xylazine.

Another advantage of using sedatives and analgesics as pre-anesthetic medication is that this produces a balanced anesthesia where the dosage of each component can be reduced, while inducing general anesthesia of sufficient depth with fewer side effects. For example, administration with midazolam and fentanyl has been shown to improve sevoflurane anesthesia for moderately painful surgeries in mice (Lipiski et al., 2017). Depressant effects of sevoflurane on the respiratory rate and the negative post-anesthetic effects on heart rate were decreased by use of pre-anesthetic medication.

Hence, to avoid unnecessary animal distress as well as to minimize potential negative effects of anesthesia on the research subject, balanced anesthesia should be the norm.

4.3 Inadequate analgesia

Lack of preventive analgesia

When anesthetics with no or minimal analgesic properties were used, e.g., in the case of isoflurane, 25% of mice and 28% of rats received no pain relief, and 13% of mice and 18% of rats received pain relief only at the end of surgery. Anesthetic agents without analgesic properties such as isoflurane and sevoflurane should always be combined with analgesics to ensure sufficient pain relief at the time the animal regains consciousness (Flecknell, 2016; Miller and Richardson, 2011). Since most analgesics only reach full onset of action 15 to 45 minutes after administration, it is advisable to administer the analgesic preoperatively. Preventive analgesia reduces both noxious stimuli reaching the central nervous system during surgery and peripheral inflammation (Miller and Richardson, 2011). Perioperative analgesics may also have an anesthetic sparing effect (Penderis and Franklin, 2005).

Lack of use of local anesthetics
Splash blocks (instillation of local anesthetics at surgery site) were used in 22% of craniotomies; for 10% of mice and rat thoracotomies local anesthetics were used for intercostal infiltration or as splash blocks. Incorporating local anesthetics into intraoperative regimens of surgeries known to cause severe pain is a valuable refinement. Combining lidocaine (10 mg/kg) and bupivacaine (5 mg/kg) provides safe and effective analgesia in a range of species (Flecknell, 2018).

Absence of multimodal analgesia
Multimodal analgesia, the use of different classes of analgesics, including local anesthetics, opioids, and nonsteroidal anti-inflammatory drugs (NSAIDs) in combination, represents current best practice. Besides this approach being the most effective in managing pain, it also helps to avoid side effects caused by the use of high doses of a single analgesic agent (Miller and Richardson, 2011; Flecknell, 2016). However, in the reviewed studies, multimodal analgesia was rarely chosen.

Intake via the drinking water
The sole application of an analgesic agent in the drinking water was a frequent method of analgesic administration, but effectiveness of this approach for managing pain is highly questionable as it is uncertain if the individual animal consumes sufficient analgesic to reach effective plasma concentrations (Sauer et al., 2016). This is especially difficult to achieve during the light phase of the animal's photoperiod when mice and rats drink infrequently (Graf et al., 2016). An additional problem can be the unfamiliar taste of the medicated drinking water. To accustom the animals to the water mix, provision should start a day or two before the surgical intervention (as planned in one of the reviewed proposals). A recent study demonstrated that the mixture of tramadol and paracetamol was voluntarily consumed via the drinking water in sufficient amounts to alleviate mild to moderate postsurgical pain (Jirkof et al., 2018). It was concluded that this protocol should at least partially alleviate pain, but the authors considered that additional analgesic agents would be required after surgical interventions that caused more severe pain. Analgesic delivery via the drinking water should be combined with another analgesic, given by injection, to provide analgesia during the light phase of the animal's photoperiod when water intake is decreased (Sauer et al., 2016). It is also practicable to train animals to consume analgesics in a palatable base, and to dose animals individually in this way (Kalliokoski et al., 2011; Leach at al., 2010).

Studies on the efficacy of the oral application of metamizole, an agent that seems especially popular in German laboratories (see Fig. 3 and 6), appear lacking. Metamizole is a pyrazolone derivate with antipyretic and analgesic properties. Its use in medical practice has been prohibited in many countries since the 1970s after its use was linked to agranulocytosis. However, it appears that the risk of metamizole-induced agranulocytosis was exaggerated (Jasiecka et al., 2014) and the agent continues to be available in Germany. A study to verify if it is an effective postoperative analgesic in rodents is urgently needed.

Analgesic dosing
Analgesic doses should be selected according to the anticipated and then actual severity of the surgical intervention, the strain and the animal's individual response to the treatment (Wright-Williams et al., 2013). The general dose recommendations for carprofen given to mice are 4-5 mg/kg (e.g., GV-SOLAS and TVT, 2015) and for ketoprofen 2-5 mg/kg (Matsumiya et al., 2012). Depending on the degree of acute pain, a 2- to 4-fold higher dose can be needed for the analgesic to be efficient (Matsumiya et al., 2012). However, these increased doses of NSAIDs may result in undesirable side effects such as gastrointestinal ulceration.

Meloxicam, at the doses proposed, is unlikely to be effective in managing postoperative pain since significantly higher doses are required (Roughan et al., 2016; Wright-Williams et al., 2007). A recent re-assessment of the efficacy of meloxicam for post-laparotomy pain in mice suggested that it had anti-inflammatory activity but failed to prevent pain, even at high doses (Roughan et al., 2016). A slow-release formulation of meloxicam also did not control pain of rats after skin incision under general anesthesia (Seymour et al., 2016).

Slow release formulations
Repeated handling and restraint causes distress (Balcombe et al., 2004; Meijer et al., 2006) and can add to postsurgical pain in newly operated animals (Jirkof et al., 2015). Thus, the frequency of analgesic injections should be minimized. However, the serum concentrations for, e.g., buprenorphine (Temgesic®) indicate that its duration of action in mice is less than 8 hours (Jirkof et al., 2015). Thus, to maintain effective serum concentrations, this agent should be given at approximately 6-hour intervals. Slow-release formulations could be administered to reduce this to once every 24-48 hours in mice (Jirkof et al., 2015; Kendall, 2016) and up to once every 72 hours in rats (Foley et al., 2011; Chum, 2014). However, these slow-release formulations are currently only available in the USA.

Non-provision of analgesia: reasons, responsibilities, and refinement
Approximately 30% of surgeries on mice and rats were not to be followed with any postoperative analgesia. However, pain occurrence is to be expected for all surgical procedures. Annex VIII of Directive 2010/63/EU (similar to previous guidance documents on severity of procedures) classifies "*surgery under general anaesthesia and appropriate analgesia, associated with post-surgical pain, suffering or impairment of general condition*" as moderate in severity and "*surgical and other interventions in animals under general anaesthesia which are expected to result in severe or persistent moderate postoperative pain, suffering or distress or severe and persistent impairment of the general condition of the animals*" as severe (EU, 2010). In contrast, almost 40% of the 684 surgeries included in our review were classified as mild by the researchers (Herrmann and Flecknell, 2018b). Explanations for not expecting and not detecting pain might be the overall low frequency of observation of the animals, the absence of structured pain assessments, as well as a general lack of knowledge about pain-related behaviors in mice and rats and the impact/severity of surgical interventions performed on them.

Ensuring that the animals used in procedures experience the minimum of pain, distress, and suffering is one of the investigator's primary responsibilities (Germany, 1986). Research applications for surgical procedures not planning to provide postoperative pain management with no scientifically plausible explanation should not be approved. If withholding pain relief is indeed necessary from a scientific standpoint, the harm-benefit analysis has to demonstrate that the anticipated benefit of the research outweighs and thus justifies the painful experience for the animal. When analyzing the restrictions of research studies in detail, a solution/compromise can often be found. It might not be possible to provide the animal with the best available pain management, but it is almost always possible to improve the animal's welfare, at least by avoiding factors that can cause and increase pain and distress. Appropriate intra- and postoperative measures also help the animals to recover more rapidly.

Postsurgical pain is likely to be present for at least 48 hours after surgery, and for longer when more invasive procedures such as thoracotomies or orthopedic procedures are undertaken. Since all opioids that are currently available in the EU act for less than 8 hours, repeated administration will be required to provide effective analgesia. When used alone, particularly in mice, NSAIDs provide insufficient analgesia following major surgery. Although handling and restraint to administer additional analgesics can cause stress, prior positive reinforcement training of the animals and use of tunnel handling of mice can greatly reduce this adverse effect. In any event, the stress of handling is likely to be minimal in comparison to unalleviated post-surgical pain.

4.4 Lack of recognition and assessment of pain

For 10% of all surgical procedures, the researchers stated that analgesia would not be necessary, but would be provided if signs of pain or suffering were observed. As mentioned before, information concerning structured pain assessment was generally absent. Our findings correlate with results from a study conducted in the UK where almost none of the 28 institutions in the survey used pain assessment methods. Although their efficacy was not assessed, the British laboratories routinely provided the animals with analgesics (Hawkins, 2002).

In many cases, investigators stated that animals would be monitored "frequently" without defining what this would mean. When the frequency of health inspections was specified, most of these checks were to be done once a day. Yet, daily inspections are a minimum legal requirement for healthy animals (EC, 2007). Even before Directive 2010/63/EU was published, recovery surgical procedures were generally rated as at least moderate in severity (Martini et al., 2000). Thus, animals that have undergone surgery require much more frequent monitoring to assure that pain is recognized and managed effectively.

Recommendations

Positive reinforcement training

The degree to which pain is perceived and the way the presence of pain is expressed by animals varies greatly between species, strains, sexes, age, health status, as well as among individuals. Therefore, proper pain assessment schemes are necessary as it is insufficient to administer an analgesic and assume that pain is alleviated. Observations to assess pain are more likely to be successful when animals are familiar with the people who observe and handle them and provide veterinary care. This highlights the importance of frequent positive interactions with the animals. Besides gentle handling practices (Hurst and West, 2010; Gouveia and Hurst, 2013), successes of positive reinforcement training in mice (Leidinger, 2018) and rats (Schuppli et al., 2017) leading to positive animal-human relationships that could help with ensuring their pain is managed effectively.

Monitoring criteria

Common behavioral alterations in mice and rats that can indicate pain are, e.g., decreased activity levels (Wright-Williams et al., 2007; Karas, 2002), sleep disturbances, reduced water and food intake (Liles et al., 1993; Carstens and Moberg, 2000), changes in nest building (Deacon, 2012; Jirkof et al., 2013; Jirkof, 2014), burrowing (Deacon, 2006, 2009, 2012; Jirkof et al., 2010; Jirkof, 2014), and grooming behavior (Miller et al., 2016). Specific behaviors associated with abdominal pain in rodents include pressing the abdomen against the cage floor, raising the tail while walking, flinching, twitching of back muscles, partial loss of balance when walking, and lifting one leg straight out behind (Wright-Williams et al., 2007; Leach et al., 2012; Miller et al., 2012, 2016). The observation and scoring of these behaviors can be complemented with monitoring of burrowing and nest building behaviors as in some strains these have been shown to be useful not only for disease progression, but also as signs of pain and distress (Jirkof, 2014). A prerequisite is that baseline data is obtained for the particular strain.

Another approach to pain assessment is the use of the grimace scale for mice (Langford et al., 2010; Leach et al., 2012) and rats (Sotocinal et al., 2011). The Mouse Grimace Scale (MGS) and Rat Grimace Scale (RGS) have been used successfully to assess spontaneous, postoperative pain that can be detected for 36 to 48 hours after surgery (Matsumiya, 2012). For research purposes, the MGS and RGS are used by scoring images taken from digital video. This system also has been used by animal care staff and research workers (Matsumiya, 2012), who confirmed user friendliness and sensitivity of the method (e.g., Faller et al., 2015). Using multiple blinded observers, previously overlooked low-level pain was identified in a study where the system was used to assess the efficacy of postsurgical analgesia protocols following thoracotomy for surgical induction of myocardial infarction (Faller et al., 2015). Furthermore, this tool has shown potential in the search for effective analgesic dose ranges. The use of the MGS showed that currently recommended doses of carprofen and ketoprofen were insufficient to manage postsurgical pain, and that paracetamol was an inadequate agent at any dose (Matsumiya et al., 2012). The grimace scale should be combined with monitoring other pain-related behaviors mentioned above, water and food intake, locomotor and exploratory activity, as well as changes in nest building and burrowing behavior in order to optimize pain management strategies. Structured pain assessments take time and therefore require sufficient numbers of trained personnel. However, they are central to assuring the efficacy of analgesic treatment.

Monitoring in the nocturnal phase

Rodents are generally more active in the dark phase of their photoperiod (Wells, 2017), so other useful tools to employ are remote video-monitoring of animals in their home cage and, in certain settings, the use of automated cages that are able to detect and record rapid changes in normal behavior (Miller et al., 2011; Wright-Williams et al., 2013; Jourdan et al., 2001). Limitations of currently available automated cages are that abnormal behaviors which are associated with pain such as pressing the abdomen on the cage floor or slow contortion of abdominal flank muscles are not detected.

5 Conclusions

The findings of this first-of-its kind review of original animal research proposals confirms international trends detected in various previous appraisals (see also Bara and Joffe, 2014; Taylor, 2010; Pound and Nicol, 2018; Uhlig et al., 2015). Although reporting of analgesia and anesthesia in scientific publications is often incomplete, it is highly probable that, for example, postoperative analgesia was not provided if it is not mentioned in the publication. Our assessment shows that legal requirements to comprehensively apply available methods of refinement were not followed in the study sample. In particular, provision of pain management was often either not planned or the proposed regimens were not appropriate, and optimal techniques such as multimodal analgesia were almost never used to alleviate postoperative pain in laboratory rodents. Pain assessment tools and schemes were not utilized, and effective health monitoring seemed to be lacking. To assess whether the situation has changed since Directive 2010/63/EU came into effect in 2013, it would be important for this study to be repeated, as well as expanded to other EU Member States.

Since animal research workers are legally bound to reduce discomfort, pain, and suffering of animals to an absolute minimum, a possible explanation for the deficiencies in the research applications is that they have not received sufficient education as well as adequate on-going training in refinement methods and animal welfare.

The number of investigators who underestimated the severity of their procedures (see Herrmann and Flecknell, 2018b) also suggests there is a knowledge gap. To ensure that animal research is conducted in the most responsible and least invasive way, resulting in findings that are valid and reproducible, professional assistance and comprehensive training are required. It is important to educate researchers not only on how experiments, husbandry, and care could be refined, but also why this is central to producing high quality research data. To adhere to all the Three Rs, specialists in reduction, replacement, and alternative, animal-free approaches should also be consulted during study planning. Every research institution should ensure that these experts are available to advise and assist research workers when planning, designing, conducting, and reporting their studies; and if a research institution does not have such expertise, outside specialists should be sought. Education and training is one of the focus areas of the new Directive, and a guidance document to improve and harmonize education and training in the EU was produced by an Expert Working Group (National Competent Authorities, 2014). Incorporating the learning outcomes into the relevant training modules and ensuring that all new research workers attend formal courses to deliver this training effectively should improve the implementation of refinements. Refresher training for more experienced researchers is recommended and more formal requirements for this should also be considered.

It is critical that clear descriptions regarding the provision of adequate pain management (and all other refinements) are included in research proposals and in publications to avoid needless pain and suffering of the animals. Full disclosure in the proposals and in the publications is also essential to enable the replication of the studies and to understand the findings. Researchers build their work on the work of other researchers, mostly by using information given in publications, and regulatory authorities and bodies, such as animal research and ethics committees, Animal Welfare Bodies (AWBs), and Animal Care and Use Committees (ACUCs), make their decisions on the content presented in proposals. Hence, openness about all aspects and details is imperative.

Finally, investigators should closely follow the PREPARE (Smith et al., 2017) and ARRIVE guidelines or the Gold Standard Publication Checklist (GSPC) to provide essential data required for interpretation of their results. Funding agencies and scientific journals still do not demand full disclosure of this information (Baker et al., 2014; Carbone and Austin, 2016; Enserink, 2017; Reichlin et al., 2016). A recent randomized controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines revealed that the request to fill in an ARRIVE checklist did not lead to full compliance with the guidelines in the manuscript (Hair et al., 2018). Thus, legal enforcement of compliance with preparation and reporting checklists, in addition to comprehensive training on all checklist items, appears critical to fulfil already existing legal requirements as well as responsibilities to the public, who ultimately funds research and, as opinion polls (e.g., Clemence and Leaman, 2016) show, expects that suffering is minimized.

References

Albrecht, M., Henke, J., Tacke, S. et al. (2014). Effects of isoflurane, ketamine-xylazine and a combination of medetomidine, midazolam and fentanyl on physiological variables continuously measured by telemetry in Wistar rats. *BMC Vet Res 10*, 198. doi:10.1186/s12917-014-0198-3

Baker, D., Lidster, K., Sottomayor, A. and Amor, S. (2014). Two years later: Journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol 12*, e1001756. doi:10.3410/f.718233317.793491394

Balcombe, J. P., Barnard, N. D. and Sandusky, C. (2004). Laboratory routines cause animal stress. *Contemp Top Lab Anim Sci 43*, 42-51. http://www.ingentaconnect.com/content/aalas/jaalas/2004/00000043/00000006/art00009

Ballard, T. (2009). Intraperitoneal route of administration – How accurate is this technique. *Animal Technology and Welfare 8*, 17-18.

Bara, M. and Joffe, A. R. (2014). The ethical dimension in published animal research in critical care: The public face of science. *Crit Care 18*, R15. doi:10.1186/cc13694

Baxter, G. M., Murphy, K. L., Taylor, P. M. and Wolfensohn, S. E. (2009). Chloral hydrate is not acceptable for anesthesia or euthanasia of small animals. *Anesthesiology 111*, 209. doi:10.1097/ALN.0b013e3181a8617e

Bertrand, H. G. M. J., Sandersen, C. and Flecknell, P. A. (2018). Reported analgesic and anaesthetic administration to non-human primates undergoing experimental surgical procedure: 2010-2015. *J Med Primatol 47*, 217-225. doi:10.1111/jmp.12346

BMEL – Bundesministerium für Ernährung und Landwirtschaft (2014). Anzahl der für Versuche und andere wissenschaftliche Zwecke verwendeten Wirbeltiere. Stand 01.12.2014. http://www.bmel.de/SharedDocs/Downloads/Tier/Tierschutz/2013-TierversuchszahlenGesamt.pdf?__blob=publicationFile

BMEL (2017). Versuchstierdaten 2016. https://www.bmel.de/SharedDocs/Downloads/Tier/Tierschutz/Versuchstierdaten2016.pdf?__blob=publicationFile

Carbone, L. and Austin, J. (2016). Pain and laboratory animals: Publication practices for better data reproducibility and better animal welfare. *PLoS One 11*, e0155001. doi:10.1371/journal.pone.0155001

Carstens, E. and Moberg, G. P. (2000). Recognizing pain and distress in laboratory animals. *ILAR J 41*, 62-71. doi:10.1093/ilar.41.2.62

Chisholm, J., De Rantere, D., Fernandez, N. J. et al. (2013). Carbon dioxide, but not isoflurane, elicits ultrasonic vocalizations in female rats. *Lab Anim 47*, 324-327. doi:10.1177/0023677213493410

Chum, H. H., Jampachairsri, K., McKeon, G. P. et al. (2014). Antinociceptive effects of sustained-release buprenorphine in a model of incisional pain in rats (Rattus norvegicus). *J Am Assoc Lab Anim Sci 53*, 193-197. http://www.ingentaconnect.com/content/aalas/jaalas/2014/00000053/00000002/art00011#

Clemence, M. and Leaman, J. (2016). Public Attitudes to Animal Research in 2016. A report by Ipsos MORI for the Department for Business, Energy & Industrial Strategy, Ipsos MORI Social Research Institute. https://www.ipsos.com/sites/default/files/publication/1970-01/sri-public-attitudes-to-animal-research-2016.pdf

Clowry, G. J. and Flecknell, P. A. (2000). The successful use of fentanyl/fluanisone ('Hypnorm') as an anaesthetic for intracranial surgery in neonatal rats. *Lab Anim 34*, 260-264. doi:10.1258/002367700780384771

Coria-Avila, G. A., Gavrila, A. M., Menard, S. et al. (2007). Cecum location in rats and the implications for intraperitoneal injections. *Lab Anim (NY) 36*, 25-30. doi:10.1038/laban0707-25

Coulter, C., Flecknell, P. and Richardson, C. (2009). Reported analgesic administration to rabbits, pigs, sheep, dogs and non-human primates undergoing experimental surgical procedures. *Lab Anim 43*, 232-238. doi:10.1258/la.2008.008021

Coulter, C., Flecknell, P., Leach, M. and Richardson, C. (2011). Reported analgesic administration to rabbits undergoing experimental surgical procedures. *BMC Vet Res 7*, 12. doi:10.1186/1746-6148-7-12

Danneman, P. J. and Mandrell, T. D. (1997). Evaluation of five agents/methods for anesthesia of neonatal rats. *Lab Anim Sci 47*, 386-395.

Deacon, R. (2006). Burrowing in rodents: A sensitive method for detecting behavioural dysfunction. *Nat Protoc 1,* 118-121. doi:10.1038/nprot.2006.19

Deacon, R. (2009). Burrowing: A sensitive essay, tested in five species of laboratory rodents. *Behav Brain Res 200*, 128-133. doi:10.1016/j.bbr.2009.01.007

Deacon, R. (2012). Assessing burrowing, nest construction, and hoarding in mice. *J Vis Exp (59)*, e2607. doi:10.3791/2607

EC – European Commission (2007). 2007/526/EC Commission Recommendation of 18 June 2007 on guidelines for the accommodation and care of animals used for experimental and other scientific purposes (notified under document number C(2007) 2525). *OJ L197*, 1-89. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2007.197.01.0001.01.ENG

EC (2013). Report from the Commission to the Council and the European Parliament. Seventh Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union. COM(2013) 859 final. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2013:0859:FIN:EN:PDF

Eger, E. I. (1981). Isoflurane: A review. *Anesthesiology 55*, 559-576. doi:10.1097/00000542-198111000-00014

Enserink, M. (2017). Sloppy reporting on animal studies proves hard to change. *Science 357*, 1337-1338. doi:10.1126/science.357.6358.1337

EU – European Union (2010). 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. *OJ L276*, 33-79. http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32010L0063&from=EN

Faller, K. M., McAndrew, D. J., Schneider, J. E. and Lygate, C. A. (2015). Refinement of analgesia following thoracotomy and experimental myocardial infarction using the mouse grimace scale. *Exp Physiol 100*, 164-172. doi:10.1113/expphysiol.2014.083139

Field, K. J., White, W. J. and Lang, C. M. (1993). Anaesthetic effects of chloral hydrate, pentobarbitone and urethane in adult male rats. *Lab Anim 27*, 258-269. doi:10.1258/002367793780745471

Fitzgerald, M. (2005). The development of nociceptive circuits. *Nat Rev Neurosci 6*, 507-520. doi:10.1038/nrn1701

Flecknell, P. A. (2016). *Laboratory Animal Anaesthesia*. 4th edition. Massachusetts, USA: Academic press.

Flecknell, P. A. (2018). Rodent analgesia: Assessment and therapeutics. *Vet J 232*, 70-77. doi:10.1016/j.tvjl.2017.12.017

Fleischman, R. W., McCracken, D. and Forbes, W. (1977). Adynamic ileus in the rat induced by chloral hydrate. *Lab Anim Sci 27*, 238-243. http://europepmc.org/abstract/med/857088

Fleischmann, T., Jirkof, P., Henke, J. et al. (2016). Injection anaesthesia with fentanyl-midazolam-medetomidine in adult female mice: Importance of antagonization and perioperative care. *Lab Anim 50*, 264-274. doi:10.1177/0023677216631458

Foley, P. L., Liang H. and Crichlow, A. R. (2011). Evaluation of a sustained- release formulation of buprenorphine for analgesia in rats. *J Am Assoc Lab Anim Sci 50*, 198-204. http://www.ingentaconnect.com/content/aalas/jaalas/2011/00000050/00000002/art00007#

Gaines Das, R. and North, D. (2007). Implications of experimental

technique for analysis and interpretation of data from animal experiments: Outliers and increased variability resulting from failure of intraperitoneal injection procedures. *Lab Anim 41*, 312-320. doi:10.1258/002367707781282802

Gaynor, J. S. and Muir, W. W. (2014). *Handbook of Veterinary Pain Management-E-Book*. Missouri, USA: Elsevier Health Sciences.

Ersetzen mit:

Germany (1972). Animal Welfare Act as amended in 1972. Section 9.1, Sentence 3. https://bit.ly/2Rlhlmz

Germany (1986). Animal Welfare Act as amended in 1986. Section 9. https://bit.ly/2r8Ve7m

Gouveia, K. and Hurst, J. L. (2013). Reducing mouse anxiety during handling: Effect of experience with handling tunnels. *PLoS One 8*, e66401. doi:10.1371/journal.pone.0066401

Graf, R., Cinelli, P. and Arras, M. (2016). Morbidity scoring after abdominal surgery. *Lab Anim 50*, 453-458. doi:10.1177/0023677216675188

Graham, M. L. and Prescott, M. J. (2015). The multifactorial role of the 3Rs in shifting the harm-benefit analysis in animal models of disease. *Eur J Pharmacol 759*, 19-29. doi:10.1016/j.ejphar.2015.03.040

Grimm, K. A., Lamont, L. A., Tranquilli, W. J. et al. (eds.) (2015). *Veterinary Anesthesia and Analgesia: The Fifth Edition of Lumb* and Jones. Iowa, USA: John Wiley & Sons, Inc. doi:10.1002/9781119421375

GV-SOLAS and TVT (2015). Committee on Anaesthesia of GV-SOLAS Supported by The working group 4 of TVT (German veterinary association for animal welfare). *Pain Management in Laboratory Animals*. http://www.gv-solas.de/fileadmin/user_upload/pdf_publikation/Anaest._Analgesie/Schmerztherapie_Mai2015_e.pdf

Hair, K., Macleod, M. R., Sena, E. S. and IICARus Collaboration (2018). A randomised controlled trial of an intervention to improve compliance with the ARRIVE guidelines (IICARus). *bioRxiv*, 370874. doi:10.1101/370874

Hall, T. J., Jagher, B., Schaeublin, M. and Wiesenberg, I. (1996). The analgesic drug buprenorphine inhibits osteoclastic bone resorption in vitro, but is proinflammatory in rat adjuvant arthritis. *Inflamm Res 45*, 299-302. doi:10.1007/BF02280995

Hawkins, P. (2002). Recognizing and assessing pain, suffering and distress in laboratory animals: A survey of current practice in the UK with recommendations. *Lab Anim 36*, 378-395. doi:10.1258/002367702320389044

Hawkins, P., Armstrong, R., Boden, T. et al. (2015). Applying refinement to the use of mice and rats in rheumatoid arthritis research. *Inflammopharmacology 23*, 131-150. doi:10.1007/s1078

Herrmann, K. and Flecknell, P. A. (2018a): Application of humane endpoints and humane killing methods in animal research proposals – a retrospective review. *Altern Lab Anim 46*, in print.

Herrmann, K. and Flecknell, P. A. (2018b). Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals – A retrospective review. *Altern Lab Anim 46*, in print.

Home Office (2017). Statistics of scientific procedures on living animals, Great Britain 2016. https://www.gov.uk/government/statistics/statistics-of-scientific-procedures-on-living-animals-great-britain-2016

Hooijmans, C. R., Leenaars, M. and Ritskes-Hoitinga, M. (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. *Altern Lab Anim 38*, 167-182. http://hdl.handle.net/2066/89153

Hurst, J. L. and West, R. S. (2010). Taming anxiety in laboratory mice. *Nat Methods 7*, 825-826. doi:10.1038/nmeth.1500

Jasiecka, A., Maślanka, T. and Jaroszewski, J. J. (2014). Pharmacological characteristics of metamizole. *Pol J Vet Sci 17*, 207-214. doi:10.2478/pjvs-2014-0030

Jirkof, P., Cesarovic, N., Rettich, A. et al. (2010). Burrowing behavior as an indicator of post-laparotomy pain in mice. *Front Behav Neurosci 4*, 165. doi:10.3389/fnbeh.2010.00165

Jirkof, P., Fleischmann, T., Cesarovic, N. et al. (2013). Assessment of postsurgical distress and pain in laboratory mice by nest complexity scoring. *Lab Anim 47*, 153-161. doi:10.1177/0023677213475603

Jirkof, P. (2014). Burrowing and nest building behavior as indicators of well-being in mice. *J Neurosci Methods 234*, 139-146. doi:10.1016/j.jneumeth.2014.02.001

Jirkof, P., Tourvieille, A., Cinelli, P. and Arras, M. (2015). Buprenorphine for pain relief in mice: Repeated injections vs sustained-release depot formulation. *Lab Anim 49*, 177-187. doi:10.1177/0023677214562849

Jirkof, P. (2017). Side effects of pain and analgesia in animal experimentation. *Lab Anim 46*, 123-128. doi:10.1038/laban.1216

Jirkof, P., Arras, M. and Cesarovic, N. (2018). Tramadol: Paracetamol in drinking water for treatment of post-surgical pain in laboratory mice. *Appl Anim Behav Sci 198*, 95-100. doi:10.1016/j.applanim.2017.09.021

Jourdan, D., Ardid, D. and Eschalier, A. (2001). Automated behavioural analysis in animal pain studies. *Pharmacol Res 43*, 103-110. doi:10.1006/phrs.2000.0760

Kalliokoski, O., Jacobsen, K. R., Hau, J. and Abelson, K. S. (2011). Serum concentrations of buprenorphine after oral and parenteral administration in male mice. *Vet J 187*, 251-254. doi:10.1016/j.tvjl.2009.11.013

Karas, A. Z. (2002). Postoperative analgesia in the laboratory mouse, Mus musculus. *Lab Anim (NY) 31*, 49-52. doi:10.1038/5000175

Kendall, L. V., Wegenast, D. J., Smith, B. J. et al. (2016). Efficacy of sustained-release buprenorphine in an experimental laparotomy model in female mice. *J Am Assoc Lab Anim Sci 55*, 66-73. http://www.ingentaconnect.com/content/aalas/jaalas/2016/00000055/00000001/art00012

Kilkenny, C., Browne, W. J., Cuthill, I. C. et al. (2010). Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biol 8*, e1000412. doi:10.1371/journal.pbio.1000412

Kirkden, R. D., Niel, L., Lee, G. et al. (2008). The validity of using an approach-avoidance test to measure the strength of aversion to carbon dioxide in rats. *Appl Anim Behav Sci 114*, 216-234. doi:10.1016/j.applanim.2008.03.001

Langford, D. J., Bailey, A. L., Chanda, M. L. et al. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nat Methods 7*, 447-449. doi:10.1038/nmeth.1455

Leach, M. C., Bowell, V. A., Allan, T. F. and Morton, D. B. (2002a). Aversion to gaseous euthanasia agents in rats and mice. *Comp*

Med 52, 249-257. http://www.ingentaconnect.com/contentone/aalas/cm/2002/00000052/00000003/art00010

Leach, M. C., Bowell, V. A., Allan, T. F. and Morton, D. B. (2002b). Degrees of aversion shown by rats and mice to different concentrations of inhalational anaesthetics. *Vet Rec 150*, 808-815. doi:10.1136/vr.150.26.808

Leach, M. C., Bowell, V. A., Allan, T. F. and Morton, D. B. (2004). Measurement of aversion to determine humane methods of anaesthesia and euthanasia. *Anim Welfare 13*, S77-S86.

Leach, M. C., Forrester, A. R. and Flecknell, P. A. (2010). Influence of preferred foodstuffs on the antinociceptive effects of orally administered buprenorphine in laboratory rats. *Lab Anim 44*, 54-58. doi:10.1258/la.2009.009029

Leach, M. C., Klaus, K., Miller and A. L. et al. (2012). The assessment of post-vasectomy pain in mice using behaviour and the Mouse Grimace Scale. *PLoS One 7*, e35656. doi:10.1371/journal.pone.0035656

Leidinger, C. S. (2018). Refinement strategies in breeding and keeping of laboratory mice. (PhD thesis). Freie Universität Berlin. http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000023125/Leidinger_online.pdf?hosts=

Leung, V., Rousseau-Blass, F., Beauchamp, G. and Pang, D. S. (2018). ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS One 13*, e0197882. doi:10.1371/journal.pone.0197882

Lewis, R. E., Kunz, A. L. and Bell, R. E. (1966). Error of intraperitoneal injections in rats. *Lab Anim Care 16*, 505-509.

Lidster, K., Jefferys, J. G., Blümcke, I. et al. (2016). Opportunities for improving animal welfare in rodent models of epilepsy and seizures. *J Neurosci Methods 260*, 2-25. doi:10.1016/j.jneumeth.2015.09.007

Liles, J. H. and Flecknell, P. A. (1993). The influence of buprenorphine or bupivacaine on the post-operative effects of laparotomy and bile-duct ligation in rats. *Lab Anim 27*, 374-380. doi:10.1258/002367793780745552

Lilley, E., Armstrong, R., Clark, N. et al. (2015). Refinement of Animal models of sepsis and septic shock. *Shock 43*, 304-316. doi:10.1097/SHK.0000000000000318

Lipiski, M., Arras, M., Jirkof, P. and Cesarovic, N. (2017). Premedication with fentanyl-midazolam improves sevoflurane anesthesia for surgical intervention in laboratory mice. *Exp Biol Med (Maywood) 242*, 1287-1298. doi:10.1177/1535370217707730

Liu, S., Zhen, G., Meloni, B. P. et al. (2009). Rodent stroke model guidelines for preclinical stroke trials. *J Exp Stroke Transl Med 2*, 2-27. doi:10.6030/1939-067X-2.2.2

Makowska, I. J. and Weary, D. M. (2009). Rat aversion to induction with inhalant anaesthetics. *Appl Anim Behav Sci 119*, 229-235. doi:10.1016/j.applanim.2009.04.003

Marquardt, N. (2013). Comparison of stress reactions induced by carbon dioxide (CO_2), isoflurane and sevoflurane during the phase of induction of narcosis in mice- a contribution to the refinement of animal research. (PhD thesis). Freie Universität Berlin. http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000014349/Marquardt_online.pdf;jsessionid=95FF7144C8D7E8149422D5DA13957E28?hosts=

Martini, L., Lorenzini, R. N., Cinotti, S. et al. (2000). Evaluation of pain and stress levels of animals used in experimental research. *J Surg Res 88*, 114-119. doi:10.1006/jsre.1999.5789

Matsumiya, L. C., Sorge, R. E., Sotocinal, S. G. et al. (2012). Using the mouse grimace scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. *J Am Assoc Lab Anim Sci 51*, 42-49. http://www.ingentaconnect.com/content/aalas/jaalas/2012/00000051/00000001/art00007

Meijer, M. K., Spruijt, B. M., Van Zutphen, L. F. M. and Baumans, V. (2006). Effect of restraint and injection methods on heart rate and body temperature in mice. *Lab Anim 40*, 382-391. doi:10.1258/002367706778476370

Miller, A. L. and Richardson, C. A. (2011). Rodent analgesia. *Vet Clin North Am Exot Anim Pract 14*, 81-92. doi:10.1016/j.cvex.2010.09.004

Miller, A. L., Flecknell, P. A., Leach, M. C. and Roughan, J. V. (2011). A comparison of a manual and an automated behavioural analysis method for assessing post-operative pain in mice. *Appl Anim Behav Sci 131*, 138-144. doi:10.1016/j.applanim.2011.02.007

Miller, A. L., Wright-Williams, S. L., Flecknell, P. A. and Roughan, J. V. (2012). A comparison of abdominal and scrotal approach methods of vasectomy and the influence of analgesic treatment in laboratory mice. *Lab Anim 46*, 304-310. doi:10.1258/la.2012.012078

Miller, A. L., Kitson, G. L., Skalkoyannis, B. et al. (2016). Using the mouse grimace scale and behaviour to assess pain in CBA mice following vasectomy. *Appl Anim Behav Sci 181*, 160-165. doi:10.1016/j.applanim.2016.05.020

Muir, W. W. and Hubbell, J. A. (2014). *Handbook of Veterinary Anesthesia-E-Book*. Missouri, USA: Elsevier Health Sciences.

National Competent Authorities (2014). A working document on the development of a common education and training framework to fulfil the requirements under the Directive. http://ec.europa.eu/environment/chemicals/lab_animals/pdf/Endorsed_E-T.pdf

Niel, L. and Weary, D. M. (2007). Rats avoid exposure to carbon dioxide and argon. *Appl Anim Behav Sci 107*, 100-109. doi:10.1016/j.applanim.2006.08.002

Niel, L., Stewart, S. A. and Weary, D. M. (2008). Effect of flow rate on aversion to gradual-fill carbon dioxide exposure in rats. *Appl Anim Behav Sci 109*, 77-84. doi:10.1016/j.applanim.2007.02.004

Penderis, J. and Franklin, R. J. (2005). Effects of pre-versus post-anaesthetic buprenorphine on propofol-anaesthetized rats. *Vet Anaesth Analg 32*, 256-260. doi:10.1111/j.1467-2995.2005.00183.x

Peterson, N. C., Nunamaker, E. A. and Turner, P. V. (2017). To treat or not to treat: The effects of pain on experimental parameters. *Comp Med 67*, 469-482. http://www.ingentaconnect.com/content/aalas/cm/2017/00000067/00000006/art00002

Percie du Sert, N., Alfieri, A., Allan, S. M. et al. (2017). The IMPROVE guidelines (ischaemia models: Procedural refinements of in vivo experiments). *J Cereb Blood Flow Metabol 37*, 3488-3517. doi:10.1177/0271678X17709185

Pound, P. and Nicol, C. J. (2018). Retrospective harm benefit analysis of pre-clinical animal research for six treatment interventions.

PLoS One 13, e0193758. doi:10.1371/journal.pone.0193758

Reichlin, T. S., Vogt, L. and Würbel, H. (2016). The researchers' view of scientific rigor – Survey on the conduct and reporting of in vivo research. *PLoS One 11*, e0165999. doi:10.1371/journal.pone.0165999

Rhodes, S. A. (2009). Evaluation of hypothermia in neonatal rats, implications for animal welfare and experimental data. (PhD thesis). University of Newcastle upon Tyne.

Richardson, C. A. and Flecknell, P. A. (2005). Anaesthesia and post-operative analgesia following experimental surgery in laboratory rodents: Are we making progress? *Altern Lab Anim 33*, 119-127. https://bit.ly/2PuKgXD

Roughan, J. V., Bertrand, H. G. and Isles, H. M. (2016). Meloxicam prevents COX-2-mediated post-surgical inflammation but not pain following laparotomy in mice. *Eur J Pain 20*, 231-240. doi:10.1002/ejp.712

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. London, UK: Methuen. http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc

Sauer, M., Fleischmann, T., Lipiski, M. et al. (2016). Buprenorphine via drinking water and combined oral-injection protocols for pain relief in mice. *Appl Anim Behav Sci 185*, 103-112. doi:10.1016/j.applanim.2016.09.009

Schuppli, C., Walterhouse, A., Chew, V. et al. (2017). A better life for research animals by fostering a culture of compassion amongst researchers. *ALTEX Proc 6*, 13. http://www.altex.ch/resources/WC10_entire_issue1.pdf

Seymour, T. L., Adams, S. C., Felt, S. A. et al. (2016). Post-operative analgesia due to sustained-release buprenorphine, sustained-release meloxicam, and carprofen gel in a model of incisional pain in rats (Rattus norvegicus). *J Am Assoc Lab Anim Sci 55*, 300-305. http://www.ingentaconnect.com/content/aalas/jaalas/2016/00000055/00000003/art00008

Smiler, K. L., Stein, S., Hrapkiewicz, K. L. and Hiben, J. R. (1990). Tissue response to intramuscular and intraperitoneal injections of ketamine and xylazine in rats. *Lab Anim Sci 40*, 60-64.

Smith, A. J., Clutton, R. E., Lilley, E. et al. (2017). PREPARE: Guidelines for planning animal research and testing. *Lab Anim 52*, 135-141. doi:10.1177/0023677217724823

Sotocinal, S. G., Sorge, R. E., Zaloum, A. et al. (2011). The rat grimace scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. *Mol Pain 7*, 55. doi:10.1186/1744-8069-7-55

Stokes, E., Flecknell, P. and Richardson, C. (2009). Reported analgesic and anaesthetic administration to rodents undergoing experimental surgical procedures. *Lab Anim 43*, 149-154. doi:10.1258/la.2008.008020

Taylor, K. (2010). Reporting the implementation of the Three Rs in European primate and mouse research papers: Are we making progress? *Altern Lab Anim 38*, 495-517. https://crueltyfreeinternational.org/sites/default/files/Taylor_3Rs%20reporting_ATLA_2010.pdf

Uhlig, C., Krause, H., Koch, T. et al. (2015). Anesthesia and monitoring in small laboratory mammals used in anesthesiology, respiratory and critical care research: A systematic review on the current reporting in top-10 impact factor ranked journals. *PLoS One 10*, e0134205. doi:10.1371/journal.pone.0134205

Wells, S. (2017). Pain assessment and new innovations. *ALTEX Proc 6*, 222. http://www.altex.ch/resources/WC10_entire_issue1.pdf

Wolfensohn, S., Hawkins, P., Lilley, E. et al. (2013a). Reducing suffering in experimental autoimmune encephalomyelitis (EAE). *J Pharmacol Toxicol Methods 67*, 169-176. doi:10.1016/j.vascn.2013.01.009

Wolfensohn, S., Hawkins, P., Lilley, E. et al. (2013b). Reducing suffering in animal models and procedures involving seizures, convulsions and epilepsy. *J Pharmacol Toxicol Methods 67*, 9-15. doi:10.1016/j.vascn.2012.09.001

Wong, D., Makowska, I. J. and Weary, D. M. (2013). Rat aversion to isoflurane versus carbon dioxide. *Biol Lett 9*, 20121000. doi:10.1098/rsbl.2012.1000

Wright-Williams, S. L., Courade, J. P., Richardson, C. A. et al. (2007). Effects of vasectomy surgery and meloxicam treatment on faecal corticosterone levels and behaviour in two strains of laboratory mouse. *Pain 130*, 108-118. doi:10.1016/j.pain.2006.11.003

Wright-Williams, S., Flecknell, P. A. and Roughan, J. V. (2013). Comparative effects of vasectomy surgery and buprenorphine treatment on faecal corticosterone concentrations and behaviour assessed by manual and automated analysis methods in C57 and C3H mice. *PLoS One 8*, e75948. doi:10.1371/journal.pone.0075948

Young, A. (2006). Halothane induction results in differing behaviours compared with carbon dioxide mixed with oxygen when used as a rat euthanasia agent. *Anim Technol Welfare 5*, 49-59.

Ziemann, A. E., Allen, J. E., Dahdaleh, N. S. et al. (2009). The amygdala is a chemosensor that detects carbon dioxide and acidosis to elicit fear behavior. *Cell 139*, 1012-1021. doi:10.1016/j.cell.2009.10.029

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

We acknowledge the support of the German Research Foundation and the Open Access Publication Fund of the Freie Universität Berlin. We are very grateful to Thorsten Busse for his help in developing an electronic database for comparing the data. And we are very thankful to the German foundation SET, who financially supported this project. We also thank ZEBET (FK 1329-472) whose travel grant enabled the first author to present the project results at international conferences.

2.2 Severity classification of surgical procedures and application of health monitoring in animal research proposals – a retrospective review

Citation: Herrmann, K. and Flecknell, P.A. (2018). Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals – a retrospective review. Alternatives to Laboratory Animals, 46(5), pp. 273-289. https://doi.org/10.1177/026119291804600508

ATLA **46**, 273–289, 2018

Severity Classification of Surgical Procedures and Application of Health Monitoring Strategies in Animal Research Proposals: A Retrospective Review

Kathrin Herrmann[1a] and Paul Flecknell[2]

[1]Freie Universität Berlin, Department of Veterinary Medicine, Institute of Pharmacology and Toxicology, Berlin, Germany; [2]Newcastle University, The Medical School, Comparative Biology Centre, Newcastle upon Tyne, UK

Summary — Animal experimentation has been one of the most controversial areas of animal use, mainly due to the intentional harms inflicted upon the animals used. In an effort to reduce these harms, research on *refinement* has increased significantly over the past 20 years. However, the extent to which these efforts have helped to reduce the severity of the research procedures, and thus animal suffering, is uncertain. To provide an indication of the awareness and implementation of *refinement* methods, we reviewed the experimental techniques for 684 surgical interventions described in 506 animal research applications that had been sent to the German competent authorities for approval in 2010. In this paper, we describe and discuss the severity categorisation of the proposed surgeries and the planned health monitoring strategies. We found that the researchers frequently underestimated the levels of pain, suffering, distress and lasting harm that were to be inflicted on the animals. Furthermore, the planned health monitoring strategies were generally flawed. To ensure responsible treatment of animals and high-quality science, adequate training of research workers in recognising and alleviating animal suffering is essential.

Key words: *application of refinement, harm–benefit analysis, health monitoring, health score sheets, legal requirements, refinement, severity classification.*

Address for correspondence: *Kathrin Herrmann, Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), 615 North Wolfe Street, W7032, Baltimore, Maryland 21205, USA.*
E-mail: *kherrma1@jhu.edu*

Introduction

Animal experimentation has been one of the most controversial areas of animal use, mainly due to the intentional harms inflicted upon the animals for the sake of the proposed benefits to humans. When Russell and Burch put forward their principles of *replacement*, *reduction* and *refinement* (known as the Three Rs) of animal experimentation almost 60 years ago (1), they described ways to eradicate *inhumanity* toward animals used in science. They were certain that the humane treatment of animals contributed to good science. Today, the Three Rs concept is embedded in many animal welfare regulations around the world (2–4). The European Union (EU) has declared its goal to work toward replacing live animal use completely (Recital 10 [4]). Nonetheless, as long as animal use is not fully replaced with non-animal methods, research into the third 'R', *refinement*, remains important. The application of our ever-increasing knowledge on *refinement* methods should benefit the over 127 million animals that are used annually in research, testing and education worldwide (5). However, improvements cannot be achieved unless this increased knowledge is translated into practice.

Several studies have assessed the *refinement* methods that are described in research publications featuring the use of laboratory animals, and found the reporting quality of the methods to be very poor (e.g. 6–9). Carbone and Austin's review (9), which identified a serious deficiency in the reporting of the use of anaesthetics and analgesics in laboratory animals undergoing a variety of major surgical procedures, concluded that post-surgical pain is likely to be undertreated. To gauge the *actual* use of *refinement* methods and their quality, we reviewed the use of such methods in relation to the estimation by the researchers of the severity of the planned procedures, as well as the thoroughness of the planned health monitoring

[a]*Current address: Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), 615 North Wolfe Street, W7032, Baltimore, Maryland 21205, USA. E-mail: kherrma1@jhu.edu*

strategies, in applications requesting the authorisation of basic and applied animal research in Germany. In particular, we were interested in assessing the efforts made by researchers to avoid unnecessary animal suffering. We were able to review 506 applications, which had been submitted to the German competent authorities in 2010.

The German licensing process requires research workers to provide detailed descriptions of all procedures, including all experimental refinements planned (see Sections 8 and 9 [10]). Besides the statutory provision that animal experiments must be reduced to an indispensable minimum (Section 9 [10]), another basic requirement since 1986 is that: "Pain, suffering and harm is allowed to be inflicted upon the animals only thus far as it is indispensable for the striven for purpose" (Section 9 [10]). In addition, also since 1986, researchers are expected to give a prospective severity estimation for each procedure, in order to gauge the overall severity of their projects. The results of the prospective severity classifications are used to weigh the harms inflicted on the animals against the expected benefits of the research. The conduct of a harm–benefit analysis (HBA), by the animal researchers as well as by the competent authorities, has been a legal requirement in Germany since 1986 (Section 7, para. 3 [10]) and in all of the European Union (EU) since 2013, when *Directive 2010/63/EU* (Article 38 [4]) became effective. Today, the HBA has a central role in the process of project authorisation throughout the EU.

In this paper, we discuss the prospective severity classification of surgical procedures and the planned health monitoring strategies applied to the animals during the experiments. In a previous study, we investigated anaesthesia and analgesia regimens, to assess the appropriateness of pain management (11). Furthermore, we have also reviewed whether so-called humane endpoints were included in the proposals and have examined their suitability in preventing needless suffering, and we scrutinised methods that were described in the research applications to kill the animals (12). The overall goal of these various studies was to draw inferences from the research applications on the use of *refinement* in practice, and to determine whether improvements are needed. These are the first studies of this type to assess actual research proposals for their level of adherence to *refinement*.

Methods

In Germany, the reviewing and licensing of animal research proposals are governed by each of the 16 federal states that have so-called 'competent authorities' for this duty. The Berlin state Data Protection and Freedom of Information Commissioner (Berliner Beauftragter für Datenschutz und Informationsfreiheit) was informed of our request for data for this study and had no objections (Case No. 5612.145). For certain scientific research purposes, data storage units are allowed to pass on personal data without the consent of the persons concerned, if: a) interests that would require protection are not affected by the kind of data requested, their publication and use; or b) the public interest in the conduct of the scientific project outweighs the protection-requiring interests of the persons concerned. The data transfer must have been approved ahead of time by the highest authorities of the federal states (Appendix 3, Section 30 [13]). After consultation with the highest authorities of the federal states, 14 out of the 16 gave us access to the original basic and applied research proposals that had been granted a licence.

The study was conducted anonymously (i.e. the individual research groups were not identified in the analyses). According to the method previously described in Herrmann and Flecknell (11), 506 applications were selected by using the criteria listed below. We focused on mice and rats, as these are among the most used species in the EU (14). All the selected proposals were submitted for authorisation in 2010. They included 422 recovery surgical procedures in mice, and 262 recovery surgical procedures in rats. Thus, the research proposals that were included in this study:

— were animal research proposals submitted to the German competent authorities in 2010;
— were basic or applied research studies;
— involved the use of the mouse or the rat;
— involved surgical and recovery procedures; and
— were original proposals that were granted a project licence.

The research proposals that were excluded from this study met at least one of a range of specific criteria, namely, that they:

— were applications for the generation of genetically altered mice;
— involved government-required animal testing, e.g. toxicity testing of drugs, vaccines, pesticides;
— used species other than the mouse and the rat;
— described procedures that did not involve recovery from surgery; or
— were proposals pre-assessed or rejected by members of the competent authorities.

We analysed the experimental protocols in the original applications. All the proposals included in this review received a project licence. However, before final approval for a licence, it is expected that the competent authorities request changes to the proposed (i.e. the original) protocols. It was not feasible to follow-up the results of the original assessments made by the competent authorities, or to track any amendments made during the assessment process.

Some of the original applications lacked the required information: for example, 6% (n = 42/684) of the surgical protocols did not list their severity classification, and humane endpoints were omitted from 45% (309/684) of the applications. When these applications were submitted, the inclusion of health score sheets was not yet a requirement; their inclusion became mandatory in Germany with the transposition of *Directive 2010/63/EU*, in 2013. Consequently, health score sheets were only available for 15% (108/684) of the procedures included in the proposals, as some of the competent authority members had been requesting their use prior to it becoming a standard feature in the animal research application form, in 2013. Despite the absence of some of the information that we planned to assess, we chose to review the original proposals as submitted by the researchers, so that we could review the intended use of *refinement* methods. We postulated that this would reflect the researchers' awareness and knowledge of potential refinements, and also their understanding of the contribution of *refinement* to the generation of high-quality research data. Project licences were generally granted for up to three years, with the possibility of extension for up to five years.

The severity classification of surgical procedures

The severity of a procedure is determined by the degree of pain, suffering, distress or lasting harm expected to be experienced by the individual animal due to the procedure (Annex VIII [4]). Suffering can vary in intensity and duration, and both of these dimensions should be taken into account when assessing the procedure's severity (15). To prospectively assess and classify the severity of surgical procedures according to the categories listed in Table 1, we used the guidance given in Annex VIII of *Directive 2010/63/EU* (4) (see pp. 77–79 for classification examples), as well as two previous guidance documents: the guide published in 1995 by the Swiss Authorities (Bundesamt für Veterinärwesen; 16) and the guide of the Working Group of Berlin Animal Welfare Officers (Berliner Arbeitskreis für Tierschutzbeauftragte e.V.; 17). The latter was last updated in 2010 and is based on the Swiss guide; thus it is very similar. These two earlier documents might have been consulted by the investigators in our sample when making their estimates of severity. The classifications in all three documents outlined above are mostly comparable.

When rating the procedures, for which discordance among guidance documents existed, we also incorporated our own experience. The first author is an expert in animal welfare, ethics and law with special focus on the Three Rs, and a former member of one of Germany's competent authorities (between 2007–2016); the second author has specialist expertise in anaesthesia and analgesia in laboratory animals, and has conducted research since the 1980s on issues associated with pain and distress and its alleviation.

Our prospective severity classification was based on the maximum severity that could be expected when the surgical procedure was conducted by a technically-skilled operator, including the use of all possible refinements and optimal care during and after surgery. Thus, we:

— categorised all thoracotomies as severe, as in the Swiss (16) and Berlin (17) guidance documents. This is a deviation from Annex VIII of the Directive (4), which only rates thoracotomies as 'severe' if analgesia is inadequate;

— categorised colon ascendens stent peritonitis (CASP) and caecal ligation and puncture (CLP), both models to induce peritonitis, sepsis and septic shock, as severe. This is in line with both the Swiss guidelines, which list endotoxic shock models as 'severe' (16), and the Berlin guide,

Table 1: The severity categories as defined in *Directive 2010/63/EU* (Annex VIII)

Severity category	Description
Mild	Procedures on animals as a result of which the animals are likely to experience short-term mild pain, suffering or distress, as well as procedures with no significant impairment of the well-being or general condition of the animals.
Moderate	Procedures on animals as a result of which the animals are likely to experience short-term moderate pain, suffering or distress, or long-lasting mild pain, suffering or distress, as well as procedures that are likely to cause moderate impairment of the well-being or general condition of the animals.
Severe	Procedures on animals as a result of which the animals are likely to experience severe pain, suffering or distress, or long-lasting moderate pain, suffering or distress, as well as procedures that are likely to cause severe impairment of the well-being or general condition of the animals.

Taken from Directive 2010/63/EU *(4).*

which lists the CASP model as a 'severe' procedure (17);

— categorised transplantation of functional organs as 'severe' while expecting effective management to prevent rejection. The Berlin guide (17) also estimated such procedures to be 'severe';

— categorised nephrectomies of more than one kidney as 'severe', as in the Berlin guide (17);

— rated nerve injuries (neuropathic pain models) to induce hyperalgesia and allodynia as severe, due to the fact that the pain will not be treated, which can lead to auto-mutilation in some animals (18);

— disagreed with the Berlin and the Swiss guides that categorise as 'mild' the following procedures: vasectomies, orchiectomies, insertion of catheters in peripheral blood vessels, implantation of biomedical devices (e.g. telemetry transmitters and minipumps), and subcutaneous implantation of tumour tissue (also, we looked at whether tumour size is considered, which was the case in the Berlin guidance document; 17). Here, we followed the classification of the Directive, which rates all recovery surgical procedures as at least 'moderate' in severity (4). Annex VIII of the *Directive 2010/63/EU* (4) classifies "surgery under general anaesthesia and appropriate analgesia, associated with post-surgical pain, suffering or impairment of general condition", as 'moderate' in severity, and "surgical and other interventions in animals under general anaesthesia, which are expected to result in severe or persistent moderate postoperative pain, suffering or distress or severe and persistent impairment of the general condition of the animals", as 'severe';

— disagreed with the estimate in the Berlin guide that trephination of the skull is a 'mild' procedure (17). This is a model of traumatic brain injury (TBI), and among its effects are a breakdown of the blood–brain barrier, oedema formation and inflammation (19). Depending on the extent of the head trauma, this model is either 'moderate' or 'severe'.

There were a number of factors that were not incorporated into the severity assessment:

1. *The effects of genetic modification of mice were not taken into account.* Many of the mice used in our sample projects were genetically modified (GM). However, we did not take into account any negative impact that could be associated with the genetic alteration of these animals. Thus, even though we chose the severity category assuming that all refinements would be in place (e.g. we rated all surgical interventions as at least moderate), the effects in particular GM lines could be greater.
2. *The effects of additional procedures following one-step or two-step surgical regimen were not taken into account.* Only the cumulative effects of procedures, such as those from a two-part surgery, were considered. For example, a 5/6-nephrectomy, which is a model for chronic renal failure, entails two consecutive surgeries. We did not take into consideration additional procedures (such as nociceptive tests, behavioural tests, imaging, blood collection, radiation, etc.) that were required as part of a study protocol.
3. *The severity of sham surgeries was not rated.*
4. *Finer ranking within the broad classifications of 'moderate' or 'severe' was not attempted.* We assigned the surgical procedures (and their expected consequences) solely to these two broad categories. However, to safeguard animals by guaranteeing an adequate harm–benefit analysis, the rating of harms caused by the recovery from surgical procedures should be done in a more detailed way than is currently the case.
5. *A category for 'severe long lasting or repetitive, substantially severe pain or suffering' was not included in this assessment.* This category will be addressed in a subsequent paper in *ATLA* (12).

In summary, we solely focused on the surgical procedures and their consequences for the animals, but we did not assess the overall severity for the individual animal, which would have to consider all interventions that the animal would experience over the course of the study, and which could further increase the overall severity and severity category.

We then carried out a further assessment of a number of the projects, with the aim of determining what was actually planned in terms of the implementation of refinements. We randomly selected (by using Research Randomizer) five surgical interventions involving the use of mice and rats that the researchers had rated as 'mild' or 'moderate'. We examined their descriptions in the proposals in detail, taking into account the planned use (or absence) of "methods used to reduce or eliminate pain, suffering and distress, including refinement of […] care conditions" and "humane end-points" (see Annex VIII, Section 2: Assignment criteria [4]), as well as killing methods, and we rated the severity of the procedure under these conditions.

Health monitoring strategies

The proposals were screened for information regarding health monitoring strategies for potential adverse effects, and protocols for welfare assessment (health score sheets). Next, we randomly selected (by using Research Randomizer) four health score sheets, to assess their specificity for use with the particular procedure or experiment. In addition, we looked for schemes and tools used to identify and score pain, as well as analgesia protocols, since post-surgical pain was to be expected. The score sheet

sample was examined for completeness, i.e. besides listing the appearance, behaviour and clinical signs, we checked for monitoring frequency, type of monitoring (observation or examination) and whether instructions were given on how to proceed if an animal reached a certain score (therapeutic actions or killing).

Results

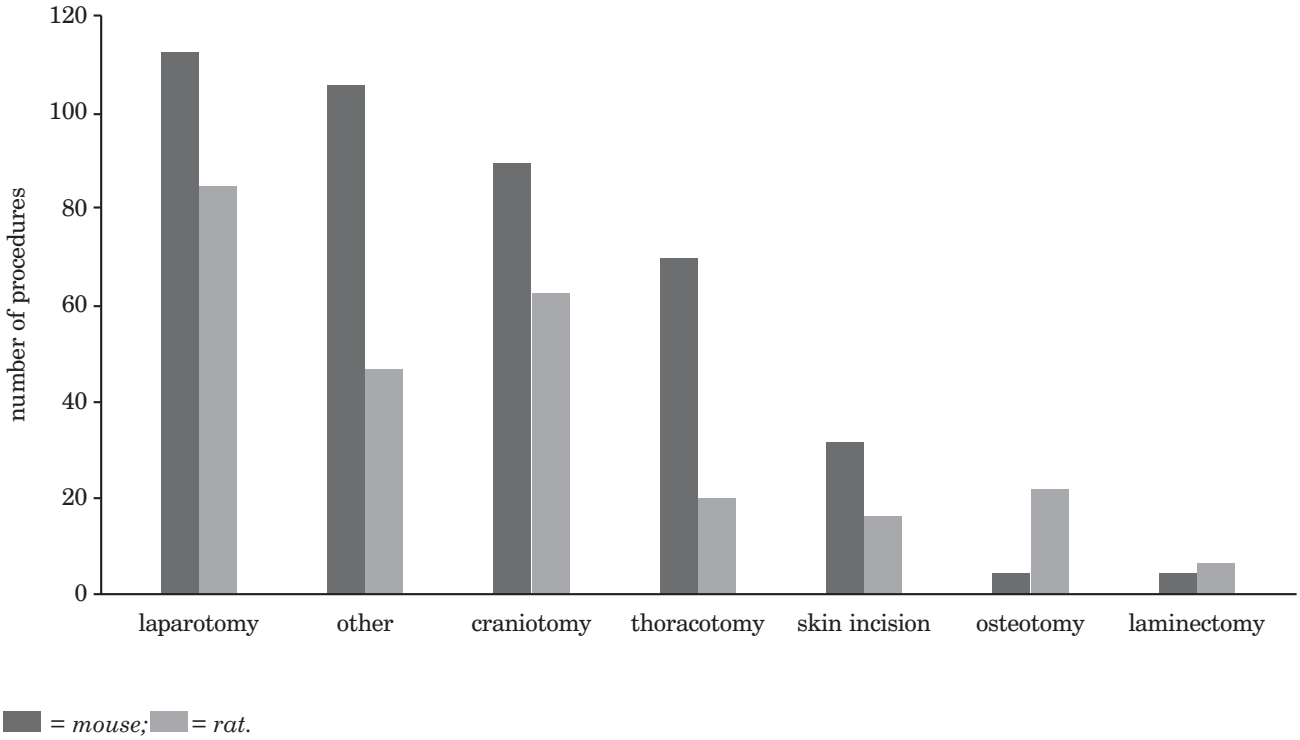
Our study sample included a large variety of recovery surgeries (Figure 1). Out of the 684 surgical procedures, 422 were performed on mice and 262 on rats. The surgical procedures performed most often on mice and rats were laparotomies, followed by craniotomies in rats. For mice, the second most common surgical procedures fell within a general category of all surgeries that did not involve the opening of a body cavity ('Other' in Figure 1); examples included neuropathic pain models, the hindlimb ischaemia model and the middle cerebral artery occlusion (MCAO) model.

Severity classification of surgical procedures

Within the assessed protocols, researchers had given estimates regarding the severity of 94% (n = 642/684) of the planned surgical procedures. We assessed the severity of the procedures, assuming best practice, i.e. that all possible *refinement* methods would be employed (as described in the protocol methods section) and compared our results (Figure 2) with the predictions given by the researchers (Figure 3).

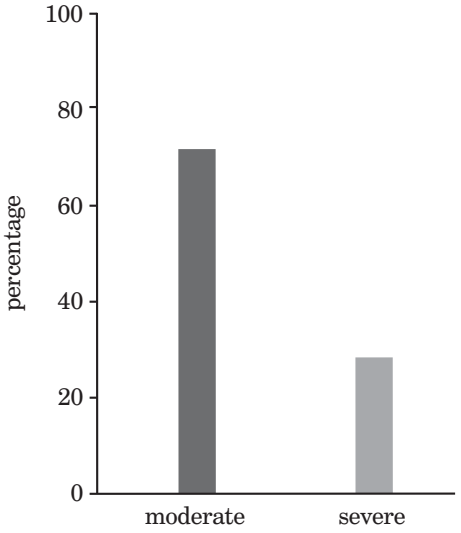
Even though all procedures were surgical procedures under general anaesthesia, almost 40% (269/684) of them were rated as being 'mild' by the researchers, whereas we classified them as either 'moderate' (246/269) or 'severe' (23/269). Figure 3 shows the highest possible severity that the researchers expected, including procedures that they had rated as 'mild to moderate' or 'moderate to severe', suggesting that the procedures might not be in the higher category. Out of 322 procedures estimated to be either 'moderate' or 'mild to moderate' by the investigators, 216 fell into the 'moderate' and 106 into the 'severe' category according to the guidance documents. The animal researchers rated 7% (51/684) of their procedures as 'severe' (including 'moderate to severe'), whereas we rated 28% (194/684) of the procedures as 'severe'. Two surgical interventions (both hindlimb ischaemia models in mice) that we rated as potentially 'moderate', were classified higher (as 'moderate to severe' and as 'severe', respectively) by the experimenters. In sum-

Figure 1: The types and numbers of surgical procedures performed on mice and rats



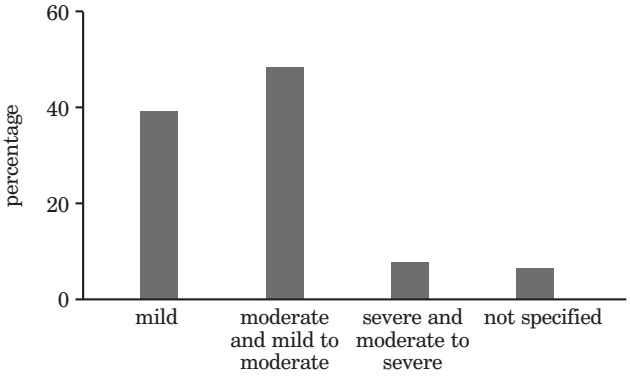
A total of 684 surgical procedures were performed, 422 on mice and 262 on rats. 'Other' refers to a general category of all surgeries that did not involve the opening of a body cavity, e.g. compression or ligation of nerves, or surgery to cause a middle cerebral artery occlusion (reproduced from Herrmann and Flecknell [11]).

Figure 2: The severity classification given by the authors to the planned surgical interventions



The severity of the procedures was assessed by the current authors according to the instructions provided in the Annex VIII of Directive 2010/63/EU *(4) and the guidance documents from Switzerland (16) and Berlin (17). For details, see* Methods *section.*

Figure 3: The severity classification given by the researchers to their planned surgical interventions



For 6% of the procedures (n = 42), severity was not classified.

mary, 59% (377/642) of the procedures rated by the researchers deviated from our prospective estimates by using the mentioned guidance documents, and in 58% (375/642), the investigators' severity classification was lower.

Next, we re-examined the five randomly selected surgical interventions and we assessed their planned implementation of refinements and the impact of this on the severity category. The examples listed in Table 2 and Table 3 include the animal model and information on anaesthesia, analgesia and other important details to gauge the severity of the procedure, and show a comparison of the severity categories as assessed by the researcher and by us, factoring in whether the researchers had included refinements, as well as their likely impact.

We estimated the effects of the surgical interventions described in Tables 2 and 3 as being 'severe', whereas the researchers rated them as 'mild' or 'moderate'. Since *refinement* methods were either not, or not fully, employed, the severity in all the examples was expected to be increased. For instance, we rated the neuropathic pain model — rated as 'mild' by the researcher — as 'severe'. This model induces mechanical allodynia, which is then exploited in an array of tests. Shortcomings comprised a lack of health score sheet use and post-operative monitoring (including pain assessment), and the absence of humane endpoints. In addition, the researchers planned to kill the animals with carbon dioxide (CO_2) in a pre-filled chamber. Aversive reactions in animals due to CO_2 inhalation had been covered extensively prior to 2010 (e.g. 20, 21); the use of pre-filled chambers is especially aversive and thus was banned as part of the implementation of the new Directive (*Directive 2010/63/EU*; see Annex IV [4]). If all possible refinements had been in place, it might have been possible to reduce the severity to 'moderate'. However, such a model could have never been rated as 'mild'.

The proposal involving myocardial infarction was rated as 'severe' by us, and as 'moderate' by the researchers. Flaws in the proposal included: the absence of local anaesthesia during surgery (intercostal infiltration); inappropriate post-operative analgesia, i.e. it should have been multimodal, combining various systemic agents with a local one; and the duration of analgesia, which might demand an extension (22). Also, the use of longer-acting agents to minimise handling of the animals in the immediate post-operative period would be an improvement. However, we did not factor this in, since slow-release formulations are currently only available in the USA. The humane endpoints in this proposal were poorly described and non-specific for the intervention. A positive feature of the application was that animals were to be assessed by a veterinarian, to evaluate whether treatment was possible. However, there should have been clearly defined endpoints in place, so that animals could be killed immediately once an endpoint was reached and thus avoid delays associated with waiting for a veterinarian to be called to examine the animal. *Directive 2010/63/EU* (Annex VIII [4]) rates thoracotomies without adequate pain management as 'severe'. If all refinements were in place, the new Directive would place this intervention in the 'moderate' category. It is, however, important to highlight that the severity

classification examples are in flux — that is, we are still at the stage of needing to continually assess the severity of many interventions, and as we acquire knowledge, many interventions are likely to change category.

The transverse aortic constriction model was classified as 'severe' by ourselves, and as 'moderate' by the researchers. The following flaws were identified: pentobarbital was used for the induction of anaesthesia, but it has a narrow safety margin, and, due to its high pH, i.p. injections can cause pain (23); there was no systemic analgesia planned for the surgery, and after surgery only "if needed" in the opinion of the researcher; and local anaesthesia was lacking, both intra-operatively and post-operatively. The applicant did not supply any information on health monitoring strategies or pain assessment, nor on the use of health score sheets or the applica-

Table 2: Examples of the use of refinements and overall severity assessments for mice undergoing recovery surgery

	Model		
	Neuropathic pain (SNI = Spared Nerve Injury model of induced mechanical allodynia)	**Myocardial infarction** (LAD = Ligation of left anterior descending artery)	**Heart hypertrophy** (TAC = Transverse Aortic Constriction model)
Species	Mouse (GM; WT)	Mouse (GM; WT)	Mouse (GM; WT)
Duration of whole experiment	4 weeks	2, 5, 7 or 21 days	14 days or 2 months
Surgery	Spared nerve injury	Thoracotomy	Thoracotomy
Endotracheal intubation	No	Yes	Yes
General anaesthesia:			
— induction	Isoflurane	Isoflurane	Pentobarbital i.p.[a]
— maintenance	Isoflurane	Isoflurane	Isoflurane
Local anaesthesia	None	None before surgery; at the end, 1% xylocaine spray on wound	None[a]
Intra-operative analgesia	Fentanyl i.p.	Buprenorphine s.c.	None
Post-operative analgesia	None	Buprenorphine s.c., twice daily for 4 days	Carprofen, "if needed"[b]
Monitoring frequency	Not stated	Daily	Not stated
Structured pain assessment	No	No	No
Health score sheets	None	None	None
Humane endpoints	None	Unclear[c]	None
Killing method	Pre-filled CO_2 chamber[d]	Cervical dislocation at end of final procedure under isoflurane anaesthesia	CO_2[d]
Researchers' severity classification	Mild	Moderate	Moderate
Our severity classification	Severe	Severe	Severe

[a]A refinement to this technique is available. [b]Inadequate analgesia (according to our assessment; for a detailed explanation see the Discussion*). [c]The proposal stated that if there were indications of pain or bad general condition (e.g. animal isolated, inactive, does not react to external stimuli, hunched back), the animal would be assessed by a veterinarian, and then treated or killed. [d]Less-inhumane methods are available.*
GM = genetically modified; i.p. = intraperitoneally; s.c. = subcutaneously; WT = wild-type.

Table 3: Examples of the use of refinements and overall severity assessments for rats undergoing recovery surgery

	Model	
	Bone infection model (titanium Kirschner wire implantation into the intramedullary canals of left femur; *Staphylococcus aureus* suspension applied into the femur)	**Acute renal failure induced by sepsis** (caecal ligation and puncture [CLP] model of sepsis)[a]
Species	Rat	Rat
Duration of whole experiment	7, 14 or 42 days	2 weeks
Surgery	Osteotomy	Laparotomy
Endotracheal intubation	No	No
General anaesthesia: **— induction** **— maintenance**	 Isoflurane Ketamine/xylazine i.p.	 Isoflurane Isoflurane
Local anaesthesia	None	None
Intra-operative analgesia	No additional analgesia	None
Post-operative analgesia	Oral paracetamol every 4 hours, duration not indicated.[b] Alternatively buprenorphine s.c. 1–2/day[c]	None[b]
Monitoring frequency	Not stated	Not stated
Structured pain assessment	No	No
Health score sheets	None	None
Humane endpoints	None	Unclear[d]
Killing method	CO_2 under general anaesthesia after final experimental procedure	Organ removal under anaesthesia by using ether[e]
Researchers' severity classification	Moderate	Moderate
Our severity classification	Severe	Severe

[a]No mention of the effects of CLP, e.g. the development of peritonitis that will cause pain. [b]Inadequate analgesia (according to our assessment). [c]Not clear when alternative would be used. [d]The investigator stated that renal failure would not be associated with pain. "Survival experiments" were planned to last for two weeks; the humane endpoints listed were: oedema, neurological disorder, refusal to eat, creatinine > 250µmol/L (blood collection from the retrobulbar venous plexus under ether anaesthesia once a week). [e]A refinement to this technique is available.
i.p. = intraperitoneally; s.c. = subcutaneously.

tion of humane endpoints. Moreover, the use of CO_2 was planned for the killing of the animals. Consequently, this procedure and its sequelae for the mice were rated as 'severe'. *Directive 2010/63/EU* (4) rates thoracotomies without adequate pain management as 'severe'. As in the other example of thoracotomy, the full application of *refinement* methods would comply with the current rating according to the new Directive.

The bone infection model in rats (Table 3) — another model that we rated as 'severe' and the experimenters rated as 'moderate' — did not include information on health monitoring strategies, pain assessment, use of health score sheets and humane endpoints. The oral administration of paracetamol every four hours causes additional distress to the animal due to the frequent handling involved. Also, the efficacy of oral paracetamol has been questioned for a long time, more-recently having been confirmed as ineffective (24). An alternative analgesic, the opioid buprenorphine, was mentioned in the proposal, but it did not specify when buprenorphine would be used as an alternative to paracetamol. If all the possible refinements

were implemented, and a better pain management protocol was proposed, then the severity classification of this procedure could be lowered, perhaps by a whole category.

The acute renal failure that is induced in rats by ligating and puncturing the caecum (the CLP model; Table 3) has several deficiencies that increase the severity of an already severe model. First, intra-operative analgesia was absent, in view of the fact that isoflurane has little to no analgesic properties; post-operative analgesia was also absent. The researcher stated that renal failure does not cause pain. However, the proposal did not mention the pain and suffering caused by the CLP model generation, such as the pain and suffering from the peritonitis that would develop, or from the overall duration of the "survival experiments". The humane endpoints given were unclear, e.g. "neurological disorder" and "refusal to eat". Additionally, the proposal mentioned the collection of blood from the retrobulbar venous plexus with ether for anaesthesia. The collection of blood from this route can be highly invasive, as it can cause retrobulbar damage and infection, and corneal damage can occur with poor technique. Mahl *et al*. (25) compared retrobulbar sinus puncture with the blood sampling technique from the sublingual vein in anaesthetised rats, and found that the collection of blood from the retrobulbar plexus caused more stress and tissue damage. Another recent study with mice and rats, which compared several blood collection methods under anaesthesia, found that this particular method led to increased anxiety in both species. The authors confirmed previous findings that retrobulbar sinus puncture is the most stressful blood sampling technique in both rats and mice (26). In addition, the use of ether should be avoided because it is stressful and irritates the mucous membranes (23).

Health monitoring strategies

None of the applications mentioned how much time would be taken to observe, assess and score the animals' health condition. When the frequency of health monitoring was specified, many proposals suggested daily checking; one proposal stated that the inspection of the animals excluded weekends. Regarding the immediate post-operative period, several applications mentioned that there would be "frequent" monitoring, but the frequency was generally not specified. For 15% (108/684) of surgical procedures, a health score sheet was in place. Only a small portion of these included information about monitoring intervals. No differentiation was made between observing animals from a distance and close inspection that involved taking them out of the cage. When the proposals did outline the use of health score sheets, but then had several treatment groups undergoing different types of surgeries (for an example, see Table 5), in most cases the score sheets were not surgery-specific.

Monitoring sheet for use with mice and rats with tumours located in the abdomen or thoracic cavity, when undergoing therapeutic experiments

The checkpoints featured on one of the more detailed health score sheets submitted are given in Table 4. The rats and mice used for this particular study received an orthotopic tumour cell implantation into the prostate gland. Laparotomy under general anaesthesia with ketamine (100mg/kg) and xylazine (5mg/kg) would be followed by carprofen (5mg/kg s.c.) at the end of the surgery, and one more dose was planned about 24 hours after surgery. It was stated in the proposal that, if pain management were to interfere with the test compounds used to treat the tumour, analgesics would have to be withheld. Further details, as well as information on how many groups of animals might be affected, were absent. Checks were planned once daily, and the animals were to be weighed once weekly — if abnormalities (score ≥ 1) were detected, then the frequency of the inspections would be increased. Soaked food pellets would be provided on the cage floor, and wounds would be treated. The same score sheet was also used for experimental groups of mice and rats who were to receive an orthotopic tumour cell implantation in the thoracic cavity.

The proposed monitoring sheet used body weight as an indicator of their well-being (see Table 4). However, since animals with tumours might develop ascites, and tumour growth might also lead to weight gain, weight checks need to be combined with body condition score (BCS) assessments, in which the spine and sacrum are palpated to determine body condition (27). The monitoring frequency should be aligned with the health condition of the animals, so when the health situation worsens (for example, an animal shows distinctly elevated breathing), monitoring is intensified accordingly. Furthermore, imaging techniques could be employed to monitor internal tumour growth and define earlier humane endpoints.

Monitoring sheet for mice undergoing middle cerebral artery occlusion or femoral artery occlusion surgery

Middle cerebral artery occlusion (MCAO) surgeries to induce focal ischaemia were conducted frequently in our sample (5% of all proposals; 24/506). The checkpoints described on the score sheet for use following a permanent occlusion of the right cerebral artery in GM mice by electrocoagulation, which entails a craniotomy and opening

Table 4: The checkpoints featured on a health monitoring sheet for use with mice and rats with tumours located in the abdomen and thoracic cavity, when undergoing therapeutic experiments

Clinical sign	Degree of appearance	Score	Date/time[a]
Body weight[b]	Normal	0	
	< 10%	1	
	10–15%	2	
	15–20%	3	
	> 20%	4	
Breathing	Normal	0	
	Slightly elevated	2	
	Distinctly elevated	3	
Mobility and reaction to stimuli	Normal	0	
	Decreased	1	
	No mobility, barely mobility after stimulus	4	
Coat	Normal	0	
	Ruffled coat	1	
	Ruffled and not clean	2	
Posture and social behaviour	Normal	0	
	Hunched back	1	
	Hunched and stilted gait	2	
	Hunched and separated from group	3	
	Apathetic	4	
Faeces	Normal	0	
	Soft	1	
	Diarrhoea	2	
Orifices	Normal	0	
	Crusted	2	
Sum[c]			

The contents of this Table were directly translated from German into English.
[a]The monitoring frequency was once a day, and this column would be completed accordingly, following each assessment. [b]The body weight was assessed once a week. [c]Instruction was given on how to proceed: from score 1 = inform principal investigator, to ≥ 4 perform humane euthanasia.

of the dura mater, are shown in Table 5. It was stated that the distal MCA segment would be electrocoagulated for 10 seconds. The same score sheet was used to assess GM mice that underwent an occlusion of their right femoral artery (FAO), another model of ischaemia. For both surgeries, it was planned that anaesthesia would be induced with an intraperitoneally administered mixture of ketamine (100mg/kg) and xylazine (10mg/kg). Intra-operative and post-operative buprenorphine would be administered (0.05mg/kg s.c.), but it was not stated when this would be administered and whether repeated doses would be given. The monitoring sheet did not provide guidance on this.

The proposal comprising two models of ischaemia in mice (Table 5) should have described the severity of the expected ischaemic insults, as this is crucial for determining monitoring frequency and the overall time taken to care for these animals. A health monitoring sheet for this type of procedure should include instructions on pain management, in addition to a method of post-surgical pain assessment to guide the analgesia protocol. Besides looking for signs of pain, a sensorimotor deficit assessment needs to be included in the score sheet, since, for example, an altered gait or weakness of the hindlimbs are to be expected. Food and water intake should also be closely monitored, and special care instructions regarding soaked food pellets provided on the cage floor should be added. Moreover, the animals need to be repeatedly checked for signs of dehydration, and instructions on the administration of fluids should be given in the score sheet. The monitoring frequency after stroke induction appeared too low. It is recommended that animals should be monitored at least every six hours within the first 48 hours of stroke induction (28). To add a note on the perioperative regimen, a local block with ropivacaine helps to minimise post-surgical pain; since it was

not stated when buprenorphine would be administered intra-operatively, analgesics should be given before the painful intervention starts (28). For a detailed review of refinements for ischaemia models, see Percie du Sert *et al.* (28).

Health score sheet for severity assessment of mice after permanent ligation of the left anterior descending artery (LAD)

Table 6 shows the checkpoints described on a health score sheet from a proposal to use GM mice to model infarction and myocardial ischaemia through permanent ligation of the left anterior descending artery (LAD). This model includes performing a thoracotomy. The use of isoflurane was planned for induction of anaesthesia, followed by endotracheal intubation and subcutaneous injection of 5mg/kg carprofen prior to opening the thorax. The post-operative pain management plan entailed eight days of daily carprofen injections (2.5mg/kg s.c.), plus daily electrolyte solution (0.5ml s.c.).

The proposal stated that all animals would be inspected daily with regard to their general condition and would be weighed. If an animal displayed signs of score 3, and in case its condition worsened toward score 2, or in case the animal lost more than 20% of body weight, the experiment would be discontinued, and the animal would be killed.

The score sheet for the experimental myocardial infarction in mice (see Table 6) did not include an assessment of the animal's breathing quality and colour of the mucous membranes, two important clinical features that may be altered by this procedure. The monitoring frequency of once per day is too low, and following thoracotomy, it is likely that the animals would need assessment every six to eight hours in the immediate post-operative period. In addition, the planned pain management is unlikely to be effective: the proposed carprofen dose is half of the generally recommended dose of 5mg/kg (22); and it was recently demonstrated that currently recommended dosages of carprofen are insufficient in managing post-surgical pain (24). During and after thoracotomy, animals should receive an opioid, such as buprenorphine, and a local anaesthesia (local block, e.g. with bupivacaine). More-recently, it has been shown that pain caused by the myocardial infarction surgery can be assessed with the Mouse Grimace Scale (MGS; 29, 30), where facial expressions are scored. Hence, in future studies of this type, or other procedures involving major surgery, the MGS could be utilised (30).

Table 5: The checkpoints featured on a health monitoring sheet for use with mice undergoing middle cerebral artery occlusion or femoral artery occlusion surgery

Clinical sign	Degree of appearance	Action
Body weight loss	Up to 5%	Daily monitoring[b]
	Up to 10%	2 × /day monitoring
	10% and more; general condition OK	Monitor weight and general condition
	10% and more, plus impaired general condition or 15% weight loss[a]	Painless killing
General condition	Healthy coat; normal social activity; weight gain	None
	Low energy	Increase monitoring to 2 × /day
	Ruffled coat	Monitoring 2 × /day plus weight check
	No grooming; abnormal posture; no social interactions; extreme weight loss	Painless killing
Wound healing	Wound suture intact	None
	Scabbed over	Increased monitoring, if necessary iodise
	Wound re-opened	Debridement, suturing
	Opened dry suture, normal general condition	Increased monitoring, if necessary iodise
	Wound infected; abnormal general condition	Painless killing

The contents of this Table were directly translated from German into English.
[a]Small weight drops occur after surgery, but normally animals stabilise quickly; weight loss of 15% and more always indicates impaired general condition or infection; humane killing is indicated. [b]The planned general monitoring frequency was once a day.

Table 6: Details from a health monitoring sheet for use in the severity assessment of mice after permanent ligation of the left anterior descending artery (LAD)

Score	Quality	Condition of the animal
6	Very active	Strong; curious; fast
5	Active	Curious; fast; occasional breaks between activity
4	Limited activity	Reactive to human interaction; frequent activity breaks
3	Inactive	Uninterested in surroundings; rarely active; sleepy; decreased food intake
2	Lethargic	No activity; freeze; no food intake
1	Moribund	No activity; breathing issues; death anticipated

The contents of this Table were directly translated from German into English.

Health score sheet for rats undergoing lung transplantation

The clinical checkpoints described in Table 7 were outlined in a score sheet taken from a proposal where rats received a lung transplant. Anaesthesia was planned with a mixture of ketamine (75mg/kg i.p.) and xylazine (10mg/kg i.p.) for induction; buprenorphine would be given (0.05mg/kg s.c.), followed by endotracheal intubation and maintenance of anaesthesia with isoflurane. The surgery would take two hours. Post-operatively, every rat would be observed three, eight, 21 and 30 hours after surgery, and a score would be calculated (see Table 7). The experiment was planned to last 48 hours post-surgery. The animal would be killed when the total score exceeded 1. It was also stated that, in the first 24 hours after lung transplantation, the rats would receive 1.5mg/kg piritramide (15mg piritramide and 10ml glucose: 0.1ml/100g body weight) s.c.; after 24 hours, if seen as necessary by the researcher, piritramide would be

Table 7: Details from a health monitoring sheet for use with rats undergoing lung transplantation

Observation after surgery	Rat # 1				Score	Details/comments
	3h	8h	21h	30h		
Behaviour					0.0	Unremarkable; active; interested
					0.2	Quiet; moves when touched
					0.4	Immobile; uninterested; hunched
Body weight					0.0	Stable; increasing
					0.2	Loss of 5 to 10%
					0.4	Loss > 10%
Appearance					0.0	Smooth, shiny coat
					0.2	Dull, ruffled coat
					0.4	Very dull and ruffled coat
Temperature					0.0	Up to 1°C
					0.2	> 1°C and < 2°C
					0.4	> 2°C
Mobility					0.0	Normal
					0.2	Reduced movement; apparent pain
					0.4	Does not move; slogs along
Eating behaviour					0.0	Normal
					0.2	Does not eat, but drinks
					0.4	Does not eat or drink, dehydrated
Total score						

The contents of this Table were directly translated from German into English.
The following details were also requested on the score sheet, for recording purposes: Rat strain; Transgenic (Y/N); Date and time of surgery.

administered via the drinking water (230ml water + 4mg piritramide + 20ml 5% glucose).

The pain management protocol for the lung transplantation procedure (Table 7) seems unlikely to be adequate, and could be improved by the addition of a local anaesthetic nerve block. Post-surgical pain assessment by using facial expressions could be employed in the future. The score sheet did not include instructions on pain management. Furthermore, it is unclear which temperature (e.g. rectal?) would be monitored, and how. Water consumption by animals following surgery is frequently reduced, and in this study water intake was not assessed on the score sheet. Thus, the administration of analgesics via the drinking water is unlikely to be effective, especially since severe pain can be anticipated following thoracotomy. Moreover, there are no data available on the pharmacokinetics of piritramide after oral administration in rodents. A slow release injectable opioid would be more likely to be effective. It is unclear why animals were only monitored for 30 hours after surgery, while the duration of the experiment was 48 hours after surgery. It is also unclear why there was a gap of 13 hours, where the animals were not monitored (between eight hours and 21 hours after surgery). A surgical intervention of such severity is likely to require monitoring at least every four to six hours.

Discussion

Some proposals were lacking the required information — for example, the description of the severity of procedures was absent for 6% (n = 42/684) of the surgical procedures. However, it should be expected that researchers submit proposals containing all of the legally required information. The level of knowledge reflected in the proposals, suggests that these omissions will only be resolved by increasing the training of research workers.

The severity assessment of surgical interventions

Almost 40% (269/684) of the planned procedures were rated as 'mild' by the researchers, and more than half were rated as either 'mild' or 'mild to moderate', even though they involved surgery under general anaesthesia. Annex VIII of *Directive 2010/63/EU* classifies "surgery under general anaesthesia and appropriate analgesia, associated with post-surgical pain, suffering or impairment of general condition", as moderate in severity, and "surgical and other interventions in animals under general anaesthesia, which are expected to result in severe or persistent moderate postoperative pain, suffering or distress or severe and persistent impairment of the general condition of the animals", as severe (4). Earlier guidance documents that might have been used by the licence applicants at the time are comparable to the Directive (16, 17). The Swiss guide (16) listed a few surgical procedures, namely, non-recovery surgery and surgical interventions expected to cause minor tissue trauma (such as vasectomies, orchiectomies and i.p. application of minipumps), as 'mild', but this does not explain the researchers' underestimation of the degree of pain, suffering, distress or lasting harm. In view of the fact that, in total, more than half of the procedures were rated as potentially 'mild' by researchers, it follows that it was not only these types of procedure (i.e. those causing minor tissue trauma) that the applicants rated as 'mild', as these comprised less than 3% of all procedures. A possible explanation is that the researchers did not detect obvious signs of pain and distress in previous research studies. This may contribute to underestimates of severity, since the assessment of pain and distress in rodents is difficult. Hence, although researchers are required to have extensive experience in the field of study when applying for a project licence in Germany (a minimum of 3–5 years is generally required), it is likely that they will have received only basic training in assessing animal welfare.

Our findings with regard to pain management strategies also revealed a lack of awareness, or insufficient knowledge, to meet legal requirements. According to the *German Animal Welfare Act* (31, 32): *Any person involved in animal research and laboratory animal care must have sufficient professional knowledge*. In a previous study we found that: for almost 30% of surgical interventions, no postoperative analgesia was planned; in around 10% of surgical interventions, pain relief would only be administered if pain was detected; and no structured pain assessment schemes were described (11). Furthermore, the immediate impact of surgery and its consequences for the impairment of the animal's health and well-being might not have been assessed properly.

A recently-published study (33) retrospectively assessed harm to animals resulting from the experiments described in over 200 scientific publications, and rated many procedures and models as 'severe'. In several cases, this severity was due to missing refinements, such as analgesia. Efficacious pain management needs to be tailored to each research project, treatment group and individual animal. Researchers should carefully choose the best possible anaesthesia and analgesia, while also considering research outcomes. Besides there being a legal requirement to alleviate pain, and to keep distress and suffering to the minimum levels possible, uncontrolled pain can become an uncontrolled study variable, as pain has profound effects, not only on animal behaviour, but also on physiology (34, 35).

Thus, research that causes unnecessary suffering should not receive regulatory approval.

Health monitoring strategies

The information on health monitoring and assessment was incomplete. Health monitoring sheets, which are extremely useful when designed specifically for the experiment to objectively assess the animals' well-being, were included for 15% of the surgical interventions, and their quality varied. When the frequency of health checks was specified, it mostly appeared to be inappropriate.

Suggestions for improvements in the quality of health monitoring

Overall, there seems to be considerable scope for improvement in the quality of health monitoring. It is evident from the examples in Tables 4–7 that, besides including observations on overall appearance, posture, spontaneous and provoked behaviour, clinical signs and body weight, health monitoring sheets need to include instructions describing the actions that should be taken if the condition of the animals deteriorates. The first papers describing pain faces of mice (29) and rats (36) were published in 2010. Since then, this pain assessment tool has been shown to be useful, particularly for detecting mild pain that had been missed with conventional pain assessments (30). As our sample was from 2010, these rodent grimace scales were not yet included. However, the assessment of facial expressions should be a crucial part of today's health monitoring and have thus far been underutilised (37). For a recent review of this topic, see Descovich *et al.* (37).

It is a key prerequisite that all personnel in charge of the animals' well-being are sufficiently trained to conduct the health monitoring procedures. To avoid mistakes and oblivion toward tasks, detailed instructions for special care (e.g. special diet, extra fluids) and palliative treatments (e.g. analgesics, antibiotics), as well as monitoring frequency according to expected (and unexpected) adverse effects, need to be included in monitoring sheets. Adequate documentation of the assessments and treatments is imperative. If this is done, then all those with responsibilities for the care and well-being of the animals would be able to fulfil these responsibilities more effectively. For instance, to ensure that timely actions are taken when animals are unwell, clear and detailed instructions are essential, and humane endpoints should also be described in a way that ensures no misunderstanding can occur over when to kill an animal, since delays would cause avoidable suffering.

The assessments of animals should include observation of their appearance, posture and unprovoked behaviour while in their home cages. This permits a more-effective assessment, and reduces the degree of handling to a minimum. When required, animals can be removed from their cage for procedures, such as assessment of body weight and body condition score (27), and for administering fluids and analgesics. Instead of subjecting the animals to multiple injections, the use of slow-release analgesic formulations can help to significantly reduce the extent of handling and number of injections (38–40), which is especially important for animals with post-operative pain. Currently, these slow-release formulae are only available in the USA; however, this will hopefully change in the near future.

Since mice and rats are nocturnal animals, additional surveillance filming in the dark phase (the circadian cycle could be reversed in the laboratory) is highly recommended. Wells (41) found that many abnormal behaviours exhibited by rodents are missed because the animals are generally not observed when they are most active. Frequent observation is especially required if the animals have undergone a surgery categorised as severe. For example, it is recommended that animals are observed at least four times a day in the first 48 hours after stroke induction (28).

It was unclear how much time was planned for monitoring, as this was not stated in the proposals. To assess and score pain and other welfare-related problems, a thorough observation is needed where one looks for pain-related behaviours (e.g. pressing the abdomen against the cage floor or the twitching of back muscles), and other signs such as decreased activity, sometimes increased activity and aggression, reduced water and food intake and nest-building, and the use of the rat and mouse grimace scale (for a more detailed summary on pain assessment, see Herrmann and Flecknell [11]). To closely assess an animal takes a minimum of 5–10 minutes, depending on the severity of the procedure, and this must be repeated at least every 2–3 hours (42). This clearly has major resource implications when relatively large numbers of animals need assessment. However, the *German Animal Welfare Act* (Section 9 [10, 31]) stipulates that principal investigators, who are also the animal research licence holders, are fully responsible for ensuring that no more than the indispensable number of animals is used, that the pain, suffering and harms caused are kept to a minimum, and that staff are present in sufficient numbers and with the adequate skills to ensure proper care for the animal subjects.

The utility of health score sheets

The use of health score sheets for improved monitoring should be part of standard laboratory care, as they help achieve objective assessments of the

animals' health and well-being. Through score sheet use, assessments can be standardised to ensure consistent evaluations by different staff members. It is crucial that findings and changes to the interventions needed (such as amended analgesia protocols) are documented. This record keeping helps refine procedures and models, and can contribute to the reduction of the impact of the procedure in subsequent studies. While there are some general parameters that should always be scored, such as body weight and body condition, it is important that, in the planning phase, individual score sheets are tailored to specific experiments and treatment groups and to their expected adverse effects; these might need to be adjusted during the course of the study to fit unforeseen circumstances. Furthermore, when health score sheets are specific to the species, procedure and model, they are extremely useful tools in ensuring the recognition and appropriate alleviation of pain, and the application of humane endpoints. To maximise the utility of the score sheets, they must be properly completed each time that the animal is monitored.

Conclusion

Our review of the severity of procedures and health monitoring in proposals submitted in 2010 revealed that the levels of pain, suffering, distress and lasting harm that could be anticipated were underestimated for the majority of the planned procedures (58%). This may have contributed to the shortcomings in terms of the proposed monitoring practices, including pain alleviation (11), and to the choice of humane endpoints and killing methods (for details, see Herrmann and Flecknell [12]). These poor practices and *refinement* deficiencies need to be avoided, not only to meet legal requirements, but also because of the needless pain and suffering caused to the animals used.

Animal experimenters have to be familiar with the behaviour of the species that they are working on, in order to correctly assess their state of health and well-being. Thus, they must be skilled, educated and equipped to detect and relieve pain accordingly. Animal experimenters are not only accountable for the research they produce, but also for the humane treatment and care of the animals used in their research projects. They must also ensure that they have knowledgeable staff in sufficient numbers to assist them with these tasks (Article 24 [4]). Hence, the education and training of research workers and their staff is a prerequisite for conducting animal experiments. The extent of the flaws detected in this sample of protocols submitted in Germany places a major burden on the competent authorities. In Germany (as in many other EU Member States), competent authorities might be inadequately resourced to identify all of the deficiencies in the research applications (Dr Heidemarie Ratsch, head of the competent authority in Berlin 1996–2016, personal communication, July 2016), leading to the licensing of projects employing protocols that cause unnecessary suffering. The misestimation of harms also contributes to inaccurate prospective harm–benefit analyses. This potentially leads to licences being granted for projects in which the harms outweigh the expected benefits.

Acknowledgements

We are very grateful to Thorsten Busse for his help in developing an electronic database for comparing the data. Furthermore, we want to acknowledge the German foundation SET, Stiftung zur Förderung von Ersatz- und Ergänzungsmethoden zur Einschränkung von Tierversuchen, whose financial support made this project possible. We also thank ZEBET for the travel grant (FK 1329-472), which enabled the first author to present the project results at additional conferences. There were no competing interests, and no ethical approval was required.

References

1. Balls, M. (2009). *The Three Rs and Humanity Criterion. An Abridged Version of The Principles of Humane Experimental Technique by W.M.S. Russell and R.L. Burch*, 131pp. Nottingham, UK: FRAME.
2. Zurlo, J., Rudacille, D. & Goldberg, A.M. (1996). The Three Rs: The way forward. *Environmental Health Perspectives* **104**, 878–880.
3. Bayne, K., Ramachandra, G.S., Rivera, E.A. & Wang, J. (2015). The evolution of animal welfare and the 3Rs in Brazil, China, and India. *Journal of the American Association for Laboratory Animal Science* **54**, 181–191.
4. Anon. (2010). *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. *Official Journal of the European Union* **L276**, 20.10.2010, 33–79.
5. Knight, A. (2008). 127 million non-human vertebrates used worldwide for scientific purposes in 2005. *ATLA* **36**, 494–496.
6. Taylor, K. (2010). Reporting the implementation of the Three Rs in European primate and mouse research papers: Are we making progress? *ATLA* **38**, 495–517.
7. Bara, M. & Joffe, A.R. (2014). The ethical dimension in published animal research in critical care: The public face of science. *Critical Care* **18**, R15.
8. Uhlig, C., Krause, H., Koch, T., de Abreu, M.G. & Spieth, P.M. (2015). Anesthesia and monitoring in small laboratory mammals used in anesthesiology, respiratory and critical care research: A systematic review on the current reporting in top-10 impact factor ranked journals. *PLoS One* **10**, e0134205.
9. Carbone, L. & Austin, J. (2016). Pain and labora-

tory animals: Publication practices for better data reproducibility and better animal welfare. *PLoS One* **11**, e0155001.
10. Anon. (1986). *Deutsches Tierschutzgesetz in der Fassung von 1986*. [German Animal Welfare Act as amended in 1986] **3158**, 10.04.85. Available at: http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl186s1319.pdf (Accessed 15.10.18).
11. Herrmann, K. & Flecknell, P.A. (2018). Retrospective review of anaesthetic and analgesic regimens used in animal research proposals. *ALTEX preprint* [doi: 10.14573/altex.1804011].
12. Herrmann, K. & Flecknell, P. (2018). The application of humane endpoints and humane killing methods in animal research proposals — A retrospective review. *ATLA* [*in press*].
13. Metschke, R. & Wellbrock, R. (2002). *Datenschutz in Wissenschaft und Forschung* [Data Protection in Science and Research], 3rd revised edition, 76pp. Berlin, Germany: Berliner Beauftragter für Datenschutz und Informationsfreiheit. Available at: https://www.hu-berlin.de/de/datenschutz/einwilligung/datenschutz-in-wissenschaft-und-forschung (Accessed 15.10.18).
14. European Commission (2013). *Report from the Commission to the Council and the European Parliament. Seventh Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union, COM(2013) 859*, 14pp. Brussels, Belgium: European Parliament. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do? uri=COM:2013:0859:FIN:EN:PDF (Accessed 15.10.18).
15. Howard, B., Nevalainen, T. & Perretta, G. (2010). *The COST Manual of Laboratory Animal Care and Use: Refinement, Reduction, and Research*, 439pp. Boca Raton, London, New York: CRC Press.
16. Bundesamt für Veterinärwesen (1995). Einteilung von Tierversuchen nach Schweregraden vor Versuchsbeginn (Belastungskategorien). *Information Tierschutz 1.04*. 20pp. Available at: https://www.medizin.uni-tuebingen.de/tierschutz/116104_Belastungskatalog_schweiz_7-11_PAM.pdf (Accessed 15.10.18).
17. Arbeitskreis Berliner Tierschutzbeauftragter (2010). *Orientierungshilfe des Arbeitskreises Berliner Tierschutzbeauftragter zur Einstufung in Belastungsgrade*. Available at: http://www.gv-solas.de/file admin/user_upload/pdf_publikation/Orientierungshilfe_21-09-10.pdf (Accessed 28.08.18).
18. Colleoni, M. & Sacerdote, P. (2010). Murine models of human neuropathic pain. *Biochimica et Biophysica Acta* **1802**, 924–933.
19. Laurer, H.L. & McIntosh, T.K. (1999). Experimental models of brain trauma. *Current Opinion in Neurology* **12**, 715–721.
20. EFSA AHAW Panel (2005). Opinion of the Scientific Panel on Animal Health and Welfare on a request of the Commission related to the aspects of the biology and welfare of animals used for experimental and other scientific purposes. *EFSA Journal* **3**, 292.
21. Hawkins, P., Playle, L., Golledge, H., Leach, M., Banzett, R., Coenen, A., Cooper, J., Danneman, P., Flecknell, P., Kirkden, R., Niel, L. & Raj, M. (2006). Newcastle consensus meeting on carbon dioxide euthanasia of laboratory animals. *Animal Technology & Welfare* **5**, 1–17. Available at: https://www.nc3rs.org.uk/sites/default/files/documents/Events/First%20Newcastle%20consensus%20meeting%20report.pdf (Accessed 15.10.18).
22. Gesellschaft für Versuchstierkunde & Tierärztliche Gesellschaft für Tierschutz, Committee on Anaesthesia of GV-SOLAS supported by the Working Group 4 of TVT (German veterinary association for animal welfare) (2015). *Pain Management for Laboratory Animals*, 69pp. Available at: http://www.gv-solas.de/fileadmin/user_upload/pdf_publikation/Anaest._Analgesie/Schmerztherapie_Mai2015_e.pdf (Accessed 15.10.18).
23. Flecknell, P.A. (2016). *Laboratory Animal Anaesthesia, 4th edition*, 321pp. Cambridge, MA, USA: Academic Press.
24. Matsumiya, L.C., Sorge, R.E., Sotocinal, S.G., Tabaka, J.M., Wieskopf, J.S., Zaloum, A., King, O.D. & Mogil, J.S. (2012). Using the Mouse Grimace Scale to reevaluate the efficacy of postoperative analgesics in laboratory mice. *Journal of the American Association for Laboratory Animal Science* **51**, 42–49.
25. Mahl, A., Heining, P., Ulrich, P., Jakubowski, J., Bobadilla, M., Zeller, W., Bergmann, R., Singer, T. & Meister, L. (2000). Comparison of clinical pathology parameters with two different blood sampling techniques in rats: Retrobulbar plexus *versus* sublingual vein. *Laboratory Animals* **34**, 351–361.
26. Harikrishnan, V.S., Hansen, A.K., Abelson, K.S. & Sørensen, D.B. (2018). A comparison of various methods of blood sampling in mice and rats: Effects on animal welfare. *Laboratory Animals* **52**, 253–264.
27. Ullman-Culleré, M.H. & Foltz, C.J. (1999). Body condition scoring: A rapid and accurate method for assessing health status in mice. *Comparative Medicine* **49**, 319–323.
28. Percie du Sert, N., Alfieri, A., Allan, S.M., Carswell, H.V., Deuchar, G.A., Farr, T.D., Flecknell, P.A., Gallagher, L., Gibson, C.L., Haley, M.J., Macleod, M.R., McColl, B.W., McCabe, C., Morancho, A., Moon, L.D., O'Neill, M.J., Pérez de Puig, I., Planas, A., Ragan, C.I., Rosell, A., Roy, L.A., Ryder, K.O., Simats, A., Sena, E.S., Sutherland, B.A., Tricklebank, M.D., Trueman, R.C., Whitfield, L., Wong, R. & Macrae, I.M. (2017). The IMPROVE guidelines (Ischaemia Models: Procedural Refinements Of *In Vivo* Experiments). *Journal of Cerebral Blood Flow & Metabolism* **37**, 3488–3517.
29. Langford, D.J., Bailey, A.L., Chanda, M.L., Clarke, S.E., Drummond, T.E., Echols, S., Glick, S., Ingrao, J., Klassen-Ross, T., LaCroix-Fralish, M.L. & Matsumiya, L. (2010). Coding of facial expressions of pain in the laboratory mouse. *Nature Methods* **7**, 447.
30. Faller, K.M., McAndrew, D.J., Schneider, J.E. & Lygate, C.A. (2015). Refinement of analgesia following thoracotomy and experimental myocardial infarction using the Mouse Grimace Scale. *Experimental Physiology* **100**, 164–172.
31. Anon. (2010). *Tierschutzgesetz* in der Fassung der Bekanntmachung vom 18. Mai 2006 (BGBl. I S. 1206, 1313), das zuletzt durch Artikel 20 des Gesetzes vom 9. Dezember 2010 (BGBl. I S. 1934) geandert worden ist. [German Animal Welfare Act *as amended on 18 May 2006, last amended on 9 December 2010*].
32. Anon. (2017). *Tierschutzgesetz* in der Fassung der Bekanntmachung vom 18. Mai 2006 (BGBl. I S. 1206, 1313), das zuletzt durch Artikel 141 des Gesetzes vom 29. März 2017 (BGBl. I S. 626) geändert worden ist. [German Animal Welfare Act *as amended on 18 May 2006, last amended on 29 March*

2017]. Available at: https://www.gesetze-im-internet.de/tierschg/BJNR012770972.html (Accessed 15.10.18).
33. Pound, P. & Nicol, C.J. (2018). Retrospective harm benefit analysis of pre-clinical animal research for six treatment interventions. *PLoS One* **13**, e0193758.
34. Jirkof, P. (2017). Side effects of pain and analgesia in animal experimentation. *Lab Animal* **46**, 123–128.
35. Peterson, N.C., Nunamaker, E.A. & Turner, P.V. (2017). To treat or not to treat: The effects of pain on experimental parameters. *Comparative Medicine* **67**, 469–482.
36. Sotocinal, S.G., Sorge, R.E., Zaloum, A., Tuttle, A.H., Martin, L.J., Wieskopf, J.S., Mapplebeck, J.C., Wei, P., Zhan, S., Zhang, S. & McDougall, J.J. (2011). The Rat Grimace Scale: A partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular Pain* **7**, 55.
37. Descovich, K., Wathan, J., Leach, M.C., Buchanan-Smith, H.M., Flecknell, P., Farningham, D. & Vick, S.J. (2017). Facial expression: An under-utilised tool for the assessment of welfare in mammals. *ALTEX* **34**, 409–429.
38. Chum, H.H., Jampachairsri, K., McKeon, G.P., Yeomans, D.C., Pacharinsak, C. & Felt, S.A. (2014). Antinociceptive effects of sustained-release buprenorphine in a model of incisional pain in rats (*Rattus norvegicus*). *Journal of the American Association for Laboratory Animal Science* **53**, 193–197.
39. Jirkof, P., Tourvieille, A., Cinelli, P. & Arras, M. (2015). Buprenorphine for pain relief in mice: Repeated injections vs sustained-release depot formulation. *Laboratory Animals* **49**, 177–187.
40. Kendall, L.V., Hansen, R.J., Dorsey, K., Kang, S., Lunghofer, P.J. & Gustafson, D.L. (2014). Pharmacokinetics of sustained-release analgesics in mice. *Journal of the American Association for Laboratory Animal Science* **53**, 478–484.
41. Wells, S. (2017). Pain assessment and new innovations. [Oral presentation.] *ALTEX* **6**, *Abstracts of the 10th World Congress*, 222. Available at: http://www.altex.ch/resources/WC10_entire_issue1.pdf (Accessed 15.10.18).
42. Flecknell, P.A. (2018). Rodent analgesia: Assessment and therapeutics. *Veterinary Journal* **232**, 70–77.

2.3 Application of humane endpoints and humane killing methods in animal research proposals – a retrospective review

Citation: Herrmann, K. and Flecknell, P.A. (2018). Application of humane endpoints and humane killing methods in animal research applications – a retrospective review. Alternatives to Laboratory Animals, 46(6), pp. 317-333. https://doi.org/10.1177/026119291804600606

The Application of Humane Endpoints and Humane Killing Methods in Animal Research Proposals: A Retrospective Review

Kathrin Herrmann[1a] and Paul Flecknell[2]

[1]Freie Universität Berlin, Department of Veterinary Medicine, Institute of Pharmacology and Toxicology, Berlin, Germany; [2]Newcastle University, The Medical School, Comparative Biology Centre, Newcastle upon Tyne, UK

Summary — *Refinement* refers to the use of methods that help to minimise animal suffering in the laboratory. Research in this area has increased significantly over the past two decades. However, the extent to which refinements are applied in practice is uncertain. To provide an indication of the implementation and awareness of refinements, we reviewed the experimental techniques for 684 surgical interventions described in 506 animal research applications sent to the German competent authorities for approval in 2010. In this paper, we describe and discuss the appropriateness of the proposed humane endpoints and killing methods. We found that, when the investigators included humane endpoints in their application, these were often lacking in detail and/or were to be implemented at a late stage of suffering. In addition, the choice of method to kill the animals could be improved in the majority of the applications. We provide recommendations for future improvements, based on the recent literature. To ensure scientific rigour, avoid needless animal suffering and enable an accurate harm–benefit analysis, animal researchers have to be knowledgeable about refinement methods and apply them effectively. To assess compliance and ensure that only those studies in which potential benefits outweigh the harms are carried out, reviews such as ours — as well as retrospective assessments of actual harms and benefits — should be conducted widely and regularly, and the findings should be published.

Key words: *application of refinement, carbon dioxide (CO_2), harm–benefit analysis, humane endpoints, killing methods, legal requirements, refinement, retrospective assessment.*

Address for correspondence: *Kathrin Herrmann, Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), 615 North Wolfe Street, W7032, Baltimore, Maryland 21205, USA.*
E-mail: *kherrma1@jhu.edu*

Introduction

Until full *replacement* of animals in science is achieved, the third R of animal experimentation, *refinement*, remains of considerable importance, as it aims "to reduce to an absolute minimum the amount of distress imposed on those animals" that are still used (1). Contemporary definitions of *refinement* take Russell and Burch's first description a step further, and include "any approach which avoids or minimises the actual or potential pain, distress and other adverse effects experienced at any time during the life of the animals involved, and which enhances their well-being" (2). Fortunately, research into *refinement* methods has increased substantially since Russell and Burch introduced their Three Rs principles of *replacement*, *reduction* and *refinement*, especially in the past two decades. For instance, in the year 2000, only five research papers were found that focused on environmental enrichment, an aspect of animal housing *refinement* that helps to fulfil basic behavioural needs. In contrast, in 2016, there were around 160 papers published on environmental enrichment research for laboratory rodents alone (3). However, the animals only benefit from these efforts if the newly gained knowledge is applied in practice. Several structured and systematic literature reviews have given some insight with regard to certain experimental refinements, notably the use of anaesthetics and analgesics (e.g. 4–11), and killing methods (7, 8, 11). However, even more recent reviews of the literature cannot be relied upon, since the inclusion of animal welfare-relevant information in publications is still incomplete. Recommendations for improvement in this reporting were made in 2007 (14), and more-detailed guidance documents for reporting stan-

[a]*Current address: Johns Hopkins University, Bloomberg School of Public Health, Center for Alternatives to Animal Testing (CAAT), 615 North Wolfe Street, W7032, Baltimore, Maryland 21205, USA. E-mail: kherrma1@jhu.edu*

dards of animal-based research, such as the Gold Standard Publication Checklist (GSPC; 12) and the ARRIVE Guidelines (13), were published in 2010. However, research publications are still lacking information on the *refinement* methods employed (e.g. 9, 10).

In view of this information deficit in scientific publications, a different approach was taken in the current study. Rather than focusing on scientific publications, applications submitted to obtain licences for basic and applied animal research were reviewed, in order to assess which experimental refinements were proposed. We assessed 506 applications submitted to the German competent authorities in 2010. In Germany, researchers are required to provide comprehensive descriptions of all procedures, including all planned *refinement* methods (see Sections 8 and 9 [15]). The law requires that animals are treated in the most humane way possible, while assuring the generation of valid scientific results. When less than optimal protocols are to be applied — for example, late experimental endpoints — researchers must demonstrate that the benefits of the proposed experiments still outweigh the harms inflicted upon the animals. The harm–benefit analysis (HBA) has been a cornerstone of the German animal research regulations since 1986 (15).

For this paper, we reviewed the proposals for information on the use of humane endpoints, and assessed whether the planned humane endpoints appeared appropriate. A humane endpoint, or "less-inhumane endpoint" (16), is the "earliest indicator in an animal experiment of (potential) pain and/or distress that, within its scientific context and moral acceptability, can be used to avoid or limit adverse effects by taking actions such as humane killing, terminating the study or alleviating the pain and distress" (17). Furthermore, we appraised, from a welfare perspective, the adequacy of the methods planned for killing the animals, either at the end of the experiment or earlier in the case of unexpected suffering that warranted pre-term killing. Since 1972, the German Animal Welfare Law (Deutsches Tierschutzgesetz) has dictated that vertebrates must only be killed *after being stunned/anesthetised or, if reasonable under the circumstances, under strict avoidance of causing them pain* (Section 4, para. 1; 18). For animals used in science, the killing method of choice also requires consideration of the impact that the method could have on the research results (19). Thus, if the most humane method was not proposed, we looked at the justification of this choice by the researchers.

We have recently published the results of our review of analgesia and anaesthesia regimens (20), and our conclusions on the severity classification of surgical procedures and proposed health monitoring strategies within the sample proposals (21). The goal of this retrospective review was to critically assess the intended use of refinements in practice, determine the areas in which further progress seems necessary, and give recommendations to improve practice in future research projects. We focused on two of the most commonly used species in science: mice and rats (22). To the best of our knowledge, this is the first assessment of its kind, in that it is based on the analysis of animal research proposals instead of published animal studies.

Methods

In Germany, the assessment and authorisation of animal research proposals fall under the authority of each of the federal states who have their own 'competent authorities' for this task. The Berlin state Data Protection and Freedom of Information Commissioner (Berliner Beauftragter für Datenschutz und Informationsfreiheit) was informed of our request for data for this study and had no objections (Case Number 5612.145). For certain scientific research purposes, data storage units are allowed to pass on personal data without the consent of the persons concerned, if: a) interests that would require protection are not affected through the kind of data requested, their publication and their use; or b) the public interest in the conduct of the scientific project outweighs the protection-requiring interests of the persons concerned. The data transmission must be approved ahead of time by the highest authorities of the federal states (23). After negotiations with the highest authorities of the federal states, 14 of the 16 agreed to grant us access to basic and applied research proposals for the assessment of refinements proposed for experiments involving the use of mice and rats.

The study was conducted anonymously (i.e. the individual research groups were not identified in the analyses). According to the method previously described in Herrmann and Flecknell (20), 506 applications were selected by using the criteria listed below. All proposals were submitted for authorisation in 2010. They included 684 recovery surgical procedures (422 in mice and 262 in rats). Thus, the research proposals that were included in this study:

— were animal research proposals submitted to the German competent authorities in 2010;

— were basic or applied research studies;

— involved the use of the mouse or the rat;

— involved surgical and recovery procedures; and

— were original proposals that were granted a project licence.

The research proposals that were excluded from this study met at least one of a range of specific criteria, namely that they:

— were applications for the generation of genetically altered mice;

— involved government-required animal testing, e.g. toxicity testing of drugs, vaccines, pesticides;

— used species other than the mouse and the rat;

— described procedures that did not involve recovery from surgery; or

— were proposals pre-assessed or rejected by members of the competent authorities.

In the research proposal, experimenters have to elaborate on the expected intensity and duration of pain and suffering, as well as on any actions taken to alleviate them. Hence, besides providing information on analgesia and other means to alleviate suffering, for the case of unforeseen (as well as likely) complications that would exceed the estimated maximum amount of suffering (and thus the authorised severity category), humane endpoints must be listed. We assessed whether humane endpoints were included in the proposals and if they were clearly defined. We then selected all the research applications with surgical interventions that were rated by the investigators as 'severe', and reviewed whether they had included humane endpoints and whether they appeared specific and appropriate for the planned procedures.

We went on to look at the duration of the severity, as well as the cumulative severity caused by additional interventions performed on the same animal. However, for the scope of this study, we only assessed cumulative effects induced by surgical interventions and not the effects of any other procedures that the animals were planned to undergo.

German legislation has stipulated since 1986 that when animals have to suffer long lasting or repetitive, substantially severe pain or suffering, the investigator must give a scientific presentation as to why *the striven for research goal is presumably of outstanding importance to essential needs of humans and animals, including for the solution of scientific problems* (Section 7, para. 3; 24). Thus, when prolonged intense suffering seemed to be a feature of the research proposal, we looked for justification by the researcher in terms of scientific significance. Next, we critically appraised the methods investigators planned to use for killing the animals, both at the end of the experiment and when a humane endpoint was reached. Animal care policies in many countries, including Germany, stipulate that death must be painless, and that fear and anxiety must be minimised. Humane killing comprises the use of the least distressing and least painful methods that cause rapid loss of consciousness and subsequent death (e.g. 19, 25–27). In addition, rendering animals unconscious prior to killing is generally considered to be the most humane approach (18). For animals used in science, the killing method of choice also requires a consideration of the impact that the killing method could have on the research results, and a HBA should be carried out to find the most appropriate method under the given circumstances. Thus, we assessed whether researchers justified their choice of method and conducted a HBA, with these points in mind.

For simplicity, procedures are described here as if they were definitely conducted according to the submitted research proposal. It is important to point out that we reviewed the proposals as they were originally submitted by the investigators, since this reflected the intended application of refinements in practice, and thus presumably their level of awareness and knowledge of how to refine experimental procedures. All the proposals included in our review were granted a project licence. However, it is possible — and in many cases very likely — that the competent authority members required amendments to the proposed protocols prior to authorisation. Project licences were generally granted for up to three years, with a possibility for extension up to five years.

Results and Discussion

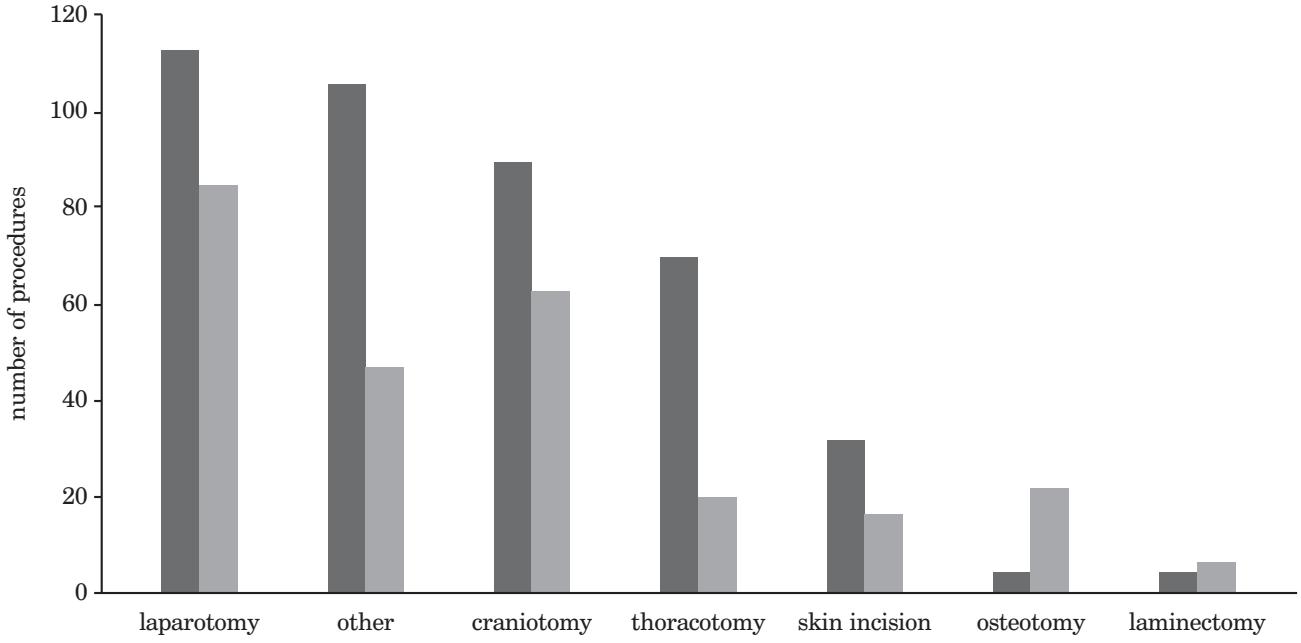
As originally submitted, some applications lacked the required information. For example, for 45% (n = 309/684) of the surgical procedures, no humane endpoints were mentioned, and for 7% (47/648) of the experiments that would end with killing (n = 648), the killing method was not specified. However, it should be expected that researchers submit proposals containing all of the legally required information.

Our study sample contained a range of recovery surgical procedures (Figure 1). Out of the 684 surgeries, 422 were performed on mice and 262 on rats. The most frequently performed surgical interventions on mice and rats were laparotomies, followed by craniotomies in rats. For mice, the second most common surgical procedures belonged to a general category of surgeries that did not involve the opening of a body cavity ('Other' in Figure 1); examples include the hindlimb ischaemia model, middle cerebral artery occlusion (MCAO) model and neuropathic pain models.

Humane endpoints

Following 55% (375/684) of surgical procedures, the researchers stated that the animals would be killed prematurely when humane endpoint(s) were reached. For the other surgical interventions, there

Figure 1: The types and numbers of surgical procedures performed on mice and rats



= *mouse;* = *rat.*

A total of 684 surgical procedures were performed, 422 on mice and 262 on rats. 'Other' refers to a general category of surgeries that did not involve the opening of a body cavity, e.g. compression or ligation of nerves, or surgery to cause a middle cerebral artery occlusion (adapted from Herrmann and Flecknell [20]).

was either no mention of humane endpoints at all, or it was stated, for example, that the experiment would be "only terminated if severe suffering occurred", "if needed, severely suffering animals would be killed" or "if complications occurred that required a veterinarian who then could not control the situation [then the animal would be killed]".

We selected all research applications in which investigators had rated surgical interventions as 'severe' and closely examined them for the inclusion of humane endpoints. Those proposals that had humane endpoints mentioned, were chosen for a detailed review in terms of the appropriateness of the endpoints. We also assessed the available information on health monitoring strategies, expected adverse effects and duration of the experiment (from surgery until the experimental endpoint was reached). In addition, when available and relevant for the discussion of the proposal, we examined the planned pain management and other care instructions. Out of 684 surgeries, 51 (7%) were rated as 'severe' by the researchers; of these 51 severe surgeries, humane endpoints were mentioned for use after 17 surgical procedures, described in 12 separate research proposals. The relevant details of these severe surgeries can be found in Table 1, and are discussed in more detail below.

Myocardial infarction models

Proposals 1 and 2 both comprise myocardial infarction models. In Proposal 1, the monitoring seemed inappropriate, especially considering the severity of the procedures. Furthermore, there was no justification for the late-stage endpoints that were proposed. The gradation of clinical signs in the health score sheet was not clearly defined, for example "dull coat, crusted orifices, dull eyes" was given a score of 2 to 3. A score of 9, which was the humane endpoint, might not have been reached until the animal had stopped eating, had lost over 20% of their body weight, had a dirty coat, uncleaned orifices, abnormal posture, dull eyes, be vocalising unusually, making noises when exhaling and auto-mutilating. In Proposal 2, the monitoring frequency was higher, but still seemed inappropriate for such a severe procedure. The long monitoring intervals might explain why the animals would be killed if they reached either the given score of III (i.e. severely altered behaviour) or IV (i.e. very severely altered behaviour).

Table 1: The humane endpoints listed in the proposals comprising surgical procedures that were rated as 'severe' by the researchers

Proposal (Species)	Name of model and type of surgical intervention(s)	Planned monitoring interval, duration of the experiment[a], additional information and humane endpoints that would lead to pre-term killing[b]
Proposal 1 (GM and WT mice)	*Experimental group*[c] *A*: Myocardial infarction model; thoracotomy; ligation of Ramus interventricularis anterior (RIVA); anticipated infarct size about 40% *Experimental group B*: Model for pressure overload-induced cardiac hypertrophy and heart failure; thoracotomy; transverse aortic constriction (TAC) *Experimental group C*: hindlimb ischaemia model, ligation of right arteria femoralis	*Monitoring*: 1st week post-surgery 1×/d; afterwards: 2×/week *Duration*: Groups A and B: 2 or 6 weeks, respectively; Group C: 3 weeks *Score sheet* comprised three main criteria that would be monitored: body weight, general condition and spontaneous behaviour. Severely affected animals (from score 9) would be killed if they reached the following conditions for the three criteria monitored: *Body weight*: food and water intake reduced; weight loss 10–15% (score 2–3) or refusal to eat; body weight loss > 20% (score 3–4) *General condition*: dull coat; crusted orifices; dull eyes (score 2–3) or dirty coat; uncleaned orifices; abnormal posture; dull eyes (score 3–4) *Spontaneous behaviour*: abnormal behaviour, reduced motor activity or hyperkinesia, not responsive, animals lie isolated from one another, no circadian rhythm (score 2–3) or unusual vocal expression; sound when exhaling; auto-amputation, deviating body care (score 3–4) *Additional humane endpoints for animals with hindlimb ischaemia*: post-surgical ulceration; necrosis or auto-amputation of underperfused limb
Proposal 2 (WT mice)	Myocardial infarction model, thoracotomy; ligation of Ramus interventricularis anterior (RIVA); ischaemia-reperfusion after 1 hour	*Duration*: 30 days *Monitoring*: for several hours, post-surgery, someone would be present; then, monitoring is planned every 12 hours *Score sheet* indicated that animal would be killed if they reached either score III or IV out of IV *Criteria for score III (passive)*: severely altered behaviour (inactive, slow movements), obvious health deficits (no food and water intake, ruffled coat, lethargy, absent body care, isolation from cage mates, altered breathing behaviour, etc.) *Criteria for score IV (lethargic)*: very severely altered behaviour (no movement or reaction to stimuli) and no food intake
Proposal 3 (WT mice)	Hypertensive heart disease model; thoracotomy; transverse aortic constriction (TAC)	Estimated failure rate of surgery (= *mortality*) for experienced surgeons: 10% *Duration*: 21 days *Monitoring/pain management*: first 5 days post-surgery every 8 hours, i.p. application of buprenorphine (0.1mg/kg); daily weighing and inspection of wound *Humane endpoints*: signs of severe impairment; weight loss (> 10% within 2 days) or eye-catching behaviour (phlegmatic, not eating/drinking, ruffled coat) or wound healing issues
Proposal 4 (WT mice)	Trauma models *Experimental group A*: polytrauma model; femur fracture, bilateral chest trauma and surrounding soft tissue trauma; blunt thoracic trauma induction by dropping a hollow cylindrical weight from a height on the chest, in addition to closed femur fracture (needle inserted into the femur canal as intramedullary pin; closure of wound; femur fracture induction using a blunt guillotine device) *Experimental group B*: monotrauma model; closed femur fracture and soft tissue trauma (see description above)	*Duration*: 0, 6, 12 and 24 hours; 3, 7, 14 and 28 days, respectively, post-surgery *Monitoring*: directly after intervention every 2 hours; after that 1×/d *Post-operative analgesia and care*: 0.8mg/ml metamizole added to drinking water for the first 3–5 days after trauma. If needed, s.c. injection with metamizole; fluid therapy: 1ml/animal 2×/d *Humane endpoints*: weight loss > 20%; wound infection; abnormal behaviour, such as lethargy or agitation, that is not explainable by external environmental factors

See end of table for full legend.

Table 1: continued

Proposal (Species)	Name of model and type of surgical intervention(s)	Planned monitoring interval, duration of the experiment[a], additional information and humane endpoints that would lead to pre-term killing[b]
Proposal 5 (Rats)	Paraplegia model to induce heterotopic ossification (HO), comprised of two surgical interventions *First surgery*: laminectomy; 2-French Fogarty catheter inserted into the dorsal epidural space through a small hole made in T10 vertebral arch, and inflated and left for 20 minutes to cause bilateral contusion *Second surgery*: osteotomy of femur; 2mm diaphyseal bone defect caused by using an osteosynthesis plate	*Duration*: 2 weeks to several months *Monitoring*: daily *Other information*: bladder (3–4×/d) and abdomen massages as functions to urinate and defecate impaired for up to 2 weeks; weight check 2×/d; fluid s.c. if needed *Post-operative analgesia*: first 3 days buprenorphine 2×/d (0.03–0.05mg/kg s.c.), followed by carprofen 2×/d (5mg/kg s.c.); if pain is detected the analgesia regimen gets prolonged *Humane endpoints*: In case of lasting severe pain (> 12 days), among others characterised by one or more signs (according to FELASA working group on pain and suffering): auto-mutilation, persistent hunched posture, vocalisation of pain while resting, complete apathy, isolation from group, obvious ruffled coat with sign of dehydration (skin stays folded), severe body weight loss (> 25% deviation from normal weight) or complete loss of appetite for 72 hours
Proposal 6 (Rats)	Hepatectomy and liver transplantation (followed by daily blood collection under isoflurane to check transplant function)	*Duration*: 7 days *Monitoring*: 3×/d *Post-operative analgesia*: tramadol via drinking water (intake monitored) *General humane endpoints*: complications during anaesthesia induction and larger surgical complications such as extended bleeding, need for resuscitation, cardiac arrest etc. *Humane endpoints during organ implantation*: vessel complications (e.g. thrombosis, leakage which leads to insufficient reperfusion of the transplant) *Humane endpoints post-surgery*: complete liver failure or failure of another vital organ
Proposal 7 (Rats)	Ischaemic stroke model (focal cerebral ischaemia); middle cerebral artery occlusion (MCAO): *First surgery*: a) Laser Doppler flow (LDF) probe and guide cannula placement on skull by using dental cement; b) permanent ligation of Arteria carotis externa and A. carotis communis; temporary blocking the origin of the MCA intraluminally with a filament *Second surgery*: 90 minutes later, 2nd anaesthesia to remove filament for reperfusion	*Mortality* due to surgery: expected to be 15% *Duration* after second surgery: 20 days *Monitoring*: daily *Other information*: functional motoric deficits expected; post-surgical analgesia not seen as necessary *Humane endpoints*: mice who have not substantially recovered from the surgeries (i.e. who do not clean themselves or show spontaneous movement, but sit apathetically in a corner) have, according to the applicant's experience, subarachnoidal bleeding and thus are killed
Proposal 8 (GM and WT mice)	Taurocholate-induced pancreatitis; model of severe necrotising pancreatitis; laparotomy *Experimental group A*: Clip ductus hepatocholedochus; incision duodenum; 1ml taurocholate in pancreatic duct; reopen clip *Experimental group B*: same procedure but no removal of clip (pancreatitis severity elevated cf. Group A); lethal experiments	*Duration, depending on experimental group*: Group A: 8 hours, 24 hours; Group B: death or 10 days; Researcher stated that animals in surgery group B would experience very severe pancreatitis with high mortality *Post-operative analgesia*: if signs of pain detected, buprenorphine (0.02mg/kg i.m.) *Monitoring*: several times a day *Humane endpoints*: surgical issues (anaesthesia, surgical technique); severely decreased food or water intake; weight loss > 20%; non-physiological, abnormal posture Group B (lethal experiments): experimental endpoint is death or 10 days after disease induction

GM = genetically modified; WT = wild type; i.m. = intramuscularly; i.p. = intraperitoneally; s.c. = subcutaneously.

See end of Table for full legend.

Table 1: continued

Proposal (Species)	Name of model and type of surgical intervention(s)	Planned monitoring interval, duration of the experiment[a], additional information and humane endpoints that would lead to pre-term killing[b]
Proposal 9 (GM and WT mice)	Sepsis models; development of endotoxic shock; laparotomy *Experimental group A*: CLP (caecal ligation and puncture) model *Experimental group B*: CASP (colon ascendens stent peritonitis) model	*Duration*: max. 96 hours (4 days) *Monitoring*: post-surgery every 4–6 hours and in advanced state every 1–2 hours *Explanation for absent post-surgical analgesia*: interference with research results due to drug metabolism *Humane endpoints*: severely altered general condition or weight loss > 20%
Proposal 10 (GM and WT mice)	Acute renal failure model; survival experiments; laparotomy *First surgery*: unilateral nephrectomy *Second surgery*: ischaemia–reperfusion injury model (IRI), clipping of A. renalis, 35 or 42 minutes, respectively	Researcher stated that > 80% of animals die within the first 4 days post-surgery *Monitoring*: for first 24 hours, there would be an hourly check; otherwise every 6 hours *Post-operative analgesia*: 2×/d buprenorphine 0.05ng/g i.m. *Humane endpoints*: weight loss > 20% or increased worsening of general condition or should pathological behaviour become apparent
Proposal 11 (GM mice)	Chronic renal failure model; 5/6-nephrectomy	*Duration*: 20 weeks Researcher stated that severe pain is only anticipated for 15 minutes referring to surgical intervention under general anaesthesia *Monitoring*: at least once a day; if injuries occurred 2–3×/d; if animal avoids movement 2×/d *Post-operative analgesia*: metamizole via drinking water over 3 days (water intake monitored) *Humane endpoints*: acute weight loss of 15% within a few days or slow weight loss of 20% because food intake is seen as objective criteria for this disease pattern; if animals feel unwell due to high concentrations of metabolites, if they become cachectic, and if constantly drinking and macerate, or if they are too weak and apathetic due to anaemia
Proposal 12 (WT mice)	Chronic renal failure model; laparotomy; 7/8-nephrectomy; 2-part surgery with 2 weeks in-between *First surgery*: 75% of right kidney gets cauterised *Second surgery*: nephrectomy left side; ALZET mini pump implantation s.c. *Third surgery*: 6 weeks later, new ALZET mini pump implantation	*Duration*: 14 weeks after first surgery *Intra-operative and post-operative analgesia:* carprofen (5mg/kg s.c.) 1×/d for 3–4 days *Score sheet* comprising five criteria that would be monitored: behaviour, weight, eating behaviour, mobility, appearance. In each category, the highest score is necessary to reach the score that equals the humane endpoint which would be: *Behaviour*: immobile, uninterested, hunched *Weight*: loss > 10% *Eating behaviour*: does not eat, does not drink, dehydrated *Mobility*: immobile, limping or dragging itself *Appearance*: dull, ruffled coat

For the purposes of this Table, the methods in the proposals were literally translated from German into English.
[a]The duration given for the experiment is from the surgery until the experimental endpoint is reached.
[b]The focus was on the effects of the described surgical procedures. However, the overall amount of pain and suffering for these animals, and hence the overall severity of the proposal, can potentially increase (or decrease) at various times during the experiment. This is because experimental surgical interventions might have detrimental effects on the animal and/or be followed by a number of other interventions (cumulative severity). Depending on the research study, these other interventions could include behavioural tests, scans, injections, blood withdrawals, gavage, irradiation, etc. The impact on severity due to genetic modification (where relevant) was not assessed.
[c]For simplicity in the current analyses, experimental groups were differentiated with A, B, or C. There were more experimental groups in the actual proposals, and thus that numbering might differ from the excerpt of the proposals described here, as only surgical experimental groups were analysed.

GM = genetically modified; WT = wild type; i.m. = intramuscularly; i.p. = intraperitoneally; s.c. = subcutaneously.

Transverse aortic constriction (TAC)

According to Proposal 3, following a transverse aortic constriction (TAC), pain relief would be given every eight hours. With regard to monitoring, it was proposed that the animal's body weight and the wound would be checked daily. It was planned that the animals would undergo a thoracotomy, which is why it would be important to monitor them at least four times a day, and this would consist mainly of observation in their home cage. Also, it was not stated when the 10% death rate would be most likely to occur. This should be noted on the health score sheet and monitoring should be intensified during this period, to ensure that these animals are humanely killed instead of dying. The humane endpoint given in the proposal needs to be clearer, as the meaning of "eye-catching behaviour" or "wound healing issues" is rather subjective. In addition, it is unclear how long the animals would be under observation to conclude that they are not eating or drinking.

Monotrauma and polytrauma models

Proposal 4 comprised a monotrauma and a polytrauma model. The proposed health monitoring was poor, since after bilateral chest trauma it can be anticipated that the mice would die. The application did not include an estimate of the anticipated mortality. However, a study on this trauma model, published in 2017, reported that 33% of the mice died within the first 30 minutes after induction of a bilateral chest contusion (28). Thus, there should be constant monitoring for at least three hours after the animals recover from anaesthesia, and, depending on the condition of the individual animals, they should be checked at least 5–6 times a day for the rest of the experiment. The applicants proposed the administration of analgesics via drinking water. Besides the fact that animals will minimise movement after such severe trauma, and that water intake is generally markedly reduced following surgery (29), it is known that water consumption follows a diurnal rhythm (30), all of which make water and thus analgesia intake too infrequent to provide adequate pain management. In addition, recent evidence shows that administration of buprenorphine via drinking water may be inadequate to reach the required plasma levels (31). Another recent study found that this route of administration seems only partially effective for animals with mild to moderate post-surgical pain (32). Injectable analgesics should be administered and soaked food should be provided on the cage floor, as normal eating and drinking ability is probably impaired. Other than the use of weight loss, the humane endpoints were unclear and thus of little benefit in avoiding unnecessary suffering.

Paraplegia and diaphyseal bone defect

Proposal 5 entailed two surgical interventions leading to paraplegia and a diaphyseal bone defect, which can be expected to lead to prolonged impairment and suffering. The rats would require intensive care and assistance with urinating and defecating. The humane endpoints in the course of the experiment appeared very late, for example, vocalisation of pain while resting and auto-mutilation. More frequent and detailed monitoring, as well as a provision for monitoring during the night, would seem imperative.

Liver transplantation

Proposal 6 comprised rats undergoing liver transplantation. The animals were to be studied for seven days after the procedure, and appeared to have required single housing, so that analgesic intake via their drinking water could be monitored. As discussed earlier, the diurnal intake of water limits the value of this route of administration, which is not efficient in providing pain relief after severe surgery. Furthermore, single housing is, in itself, stressful. The proposed humane endpoint — organ failure — does not indicate the clinical signs that will be used to assess this. The results of the daily blood sampling to monitor organ function could perhaps have been used as part of the endpoint assessment.

Ischaemia model

The humane endpoints proposed for the ischaemia model with rats in Proposal 7 could be improved and extended. The monitoring frequency appeared too low, and analgesia should be administered. A recent, comprehensive review on refinements of ischaemia models, the IMPROVE guidelines (33), provides valuable guidance which should be considered for future studies.

Severe necrotising pancreatitis

Proposal 8 describes a model involving severe necrotising pancreatitis in which high mortality is expected, especially in experimental group B. However, the proposal provides little detail of monitoring frequency, and pain management is likely to be poor in practice, since proper pain management requires frequent monitoring to assess pain. The researcher stated that mortality would be high, without giving any estimate. High mortality in the first 24 hours is explained by a systemic vascular leak, leading to hypotension and a need for fluid therapy (34). Thus, there should be continu-

ous monitoring during this period and perhaps fluid replacement therapy to reduce mortality. The researcher indicated that animals would only receive analgesia if pain was detected. Pancreatitis is highly likely to cause severe pain — in fact, this model is also used to study pain (35) — and it is known that pain from pancreatitis is particularly difficult to manage with standard analgesic agents (36). Survival experiments with the experimental endpoint of ten days after surgery, or the endpoint of death within this period, were planned for one group. Death as an endpoint should be avoided, not only from an ethical point of view, but also from a scientific standpoint (37). The researcher should have established earlier endpoints to avoid this; for example, one or more inflammatory or other molecular markers, or clinical signs, could be identified and used. This would enable fresh tissue to be collected from the animal, which could add to the value of the study (37).

Sepsis models

Proposal 9 comprised two sepsis models: caecal ligation and puncture (CLP) and colon ascendens stent peritonitis (CASP). Pain can be anticipated in this model, for example, due to peritonitis. However, the researcher stated that analgesia would not be possible, as it interfered with the model. A detailed discussion and justification of this statement were absent. The researcher might be concerned about interference with inflammatory processes, but recent studies indicate that buprenorphine can be used in rodent sepsis models (38). A general approach to dealing with this recognised issue of interactions of analgesics and research models has been suggested by Peterson and colleagues (39). The statement that the humane endpoint "severely altered general condition" is vague. In addition to the proposed endpoint of 20% body weight loss, quantitative biomarkers could perhaps be identified that predict outcome in this model. A comprehensive review on refinement of sepsis models has been provided by Lilley *et al.* (38).

Renal failure models

The acute renal failure model in the mouse, which is described in Proposal 10, has a mortality that would indicate the need for continuous monitoring for the first four days. The intramuscular route (to be used for analgesic administration) should be avoided in mice as this causes more pain than intraperitoneal or subcutaneous dosing. Endpoints other than the weight loss were unclear.

Proposals 11 and 12 concerned chronic renal failure models with late experimental endpoints (20 weeks for the 5/6-nephrectomy and 14 weeks for the 7/8-nephrectomy). Proposal 11 planned the administration of analgesics via the drinking water. As mentioned above, this is not likely to be effective. Proposal 12 stated that carprofen would be used as an analgesic, which is surprising given that non-steroidal anti-inflammatory drugs (NSAIDs) are not recommended for use in animals with impaired renal function as they might modulate the response to renal impairment (40). In Proposal 11, weight loss was the only clear endpoint — the others were non-specific and could be improved. For example, instead of stating "if they become cachectic", Body Condition Scoring (BCS; 41) should be used to set an appropriate endpoint.

Humane endpoints, if applied without delay, are seen as "a refinement strategy designed to minimise pain, suffering, or distress experienced by animals during an experiment" (42). In 1998, an international conference on the *Use of Humane Endpoints in Animal Experimentation for Biomedical Research* was held in Zeist, The Netherlands, and resulted in a special issue on the topic in the journal *Laboratory Animals* (43). Since then, progress has been made in developing more-humane endpoints, which, if implemented, would avoid unnecessary and avoidable animal suffering. Currently, the use of humane endpoints is not only a legal requirement in many countries (19), but is also generally considered best practice in animal-based research and testing (44).

Following over half of the planned surgical procedures, pre-determined endpoints were given. Thus, with adequate monitoring and pain recognition skills, it could be anticipated that unnecessary suffering could be avoided. However, the vast majority of the proposed humane endpoints were poorly described and thus unclear. In addition, they were generally not procedure-specific and were also late-stage (as seen in Table 1). To give additional examples, one proposal stated: "if it is assumed that the animal is suffering, despite the analgesia provided (absent drinking, body weight loss), the animal is killed"; in this proposal, no information on monitoring or pain assessments was provided. In another proposal, where rats would develop acute severe peritonitis and would not receive analgesia, it was stated that the animals would be killed if "the pain would exceed what is bearable". Hence, it is not only likely that such vague information would require subjective interpretation by the person monitoring the animals, but also that this uncertainty would delay the termination of the study, and lead to variability in the endpoints used. Furthermore, when monitoring intervals were actually specified, the intervals proposed often appeared too long, which could also result in avoidable pain and suffering for the animals. In order to recognise distress, pain and suffering, frequent monitoring, as well as pain recognition and assessment skills, are fundamental prerequisites.

The limits of pain and suffering should be determined based on a harm–benefit analysis, and this should incorporate a study design that maximises potential benefits and minimises anticipated harms. Most humane endpoints in the proposals that we examined could, and should, be improved. If endpoints are not clearly described with objective or quantifiable criteria (to avoid bias), the personnel responsible for monitoring the animals cannot act quickly when an animal is thought to be exceeding, or about to exceed, the agreed endpoint and thus, the severity limit allocated in the project licence. Moreover, taking into consideration the duration of the planned experiments and the potential degree of lasting substantial pain and suffering in the projects shown in Table 1, researchers should have provided an explanation as to why *the striven for research goal is presumably of outstanding importance to essential needs of humans and animals, including for the solution of scientific problems* (see Section 7, para. 3 [24]). This provision ought to ensure that projects of such severity provide an additional explanation as to why the research is justified. For one proposal, this information was not available (Proposal 5); one application ticked the box that an additional justification was provided (Proposal 9), but it was separate from the proposal and thus not available for us to assess. In the other ten applications, the researchers stated that it was "not applicable" to their proposed projects.

Two decades ago, the organisers of the first international conference on humane endpoints were sure that: "Humane endpoints are part of a dynamic process, influenced by scientific developments as well as by animal welfare concerns as they evolve with time. It is very likely, therefore, that what we consider to be 'humane' today will not be in the future, and so we need to strive constantly to develop less-inhumane endpoints, and to continually reassess and re-evaluate the endpoints in animal experiments" (45).

We urge that increased efforts are taken by the scientific community to search for earlier experimental (and thus less inhumane) endpoints. Animal subjects should also be benefitting from new technologies and other advancements widely used in humans. For example, the results of imaging could be used to determine earlier experimental endpoints. Also, biomarkers could be used to predict outcome, which, for some research settings, could be even before the onset of clinical signs. The availability of quantitative biomarkers and imaging techniques should not only advance the scientific objectives of the study but also be used to reduce the pain and distress experienced by the experimental animals.

We suggest that international working groups with expertise in the Three Rs and their respective research field are established, in order to, in a first step, review the validity of the particular animal model, and, in a second step, when proven valuable, gather and provide recommendations for the refinement of the animal model. The second part of such a critical appraisal has been conducted for a few models and research fields, such as ischaemic stroke (33), rheumatoid arthritis (46), experimental autoimmune encephalomyelitis (47), and for models and procedures involving seizures, convulsions and epilepsy (48, 49), and sepsis and septic shock (38). Such expert groups should regularly update their validity assessments and refinement recommendations, which should be seen as guidance aiding researchers who are considering and/or already employing these models.

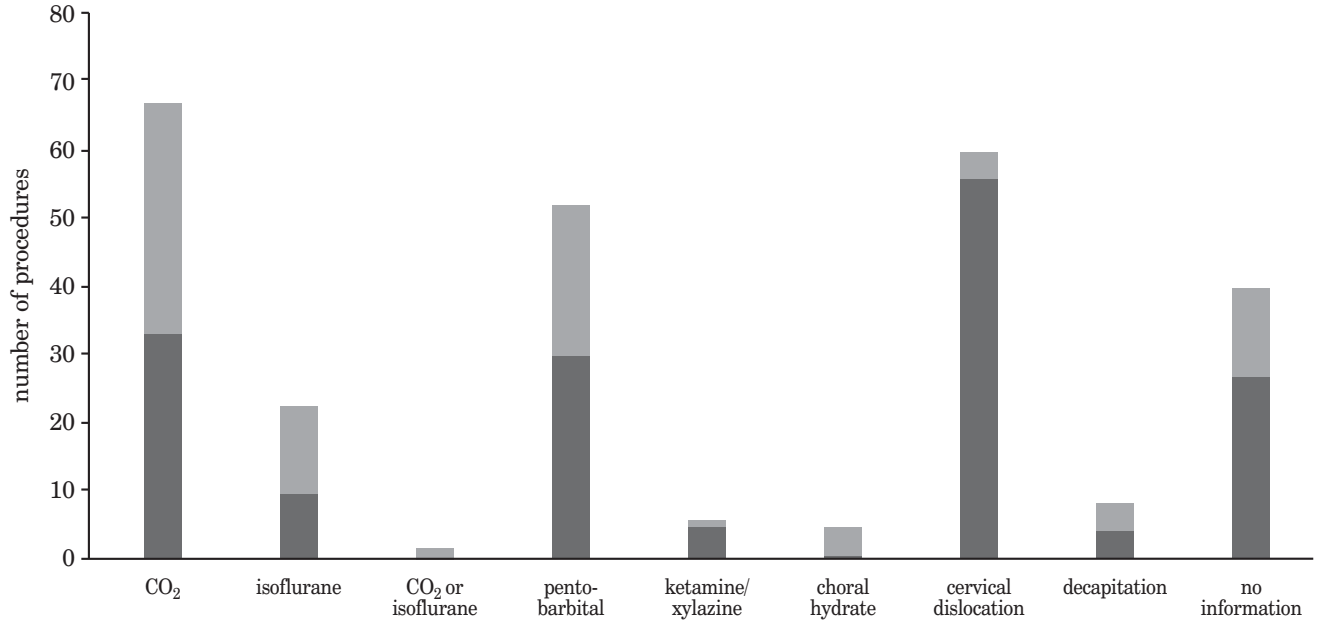
Killing methods

Anaesthetising the animal prior to killing (a two-step killing procedure) is generally considered the most humane way when ending an animal's life (18). The available data showed that, following 648 (out of 684) surgical interventions, the animals would ultimately be killed. The 36 other surgical procedures were followed by an additional recovery surgical procedure; these experiments were eventually terminated by killing as well. In 41% (264/684) of the experiments, it was planned that the animals would be killed in a one-step procedure, by using either a physical method (cervical dislocation, decapitation) or an overdose of one agent that, when given in high doses, would lead to death. The most frequently mentioned methods to terminate mouse experiments in a one-step procedure were (in decreasing order of frequency): cervical dislocation, carbon dioxide (CO_2) inhalation, and pentobarbital overdose. In conscious rats, the most commonly used means were CO_2 inhalation, an overdose of pentobarbital, and an overdose of isoflurane (Figure 2).

Out of the remaining almost 60% of experiments (384), two-thirds ended with a final procedure that required anaesthesia prior to killing, for example, for tissue perfusion or for tissue/organ collection. The animals in the other third of experiments were first anaesthetised with one agent and then killed by another means. In approximately half of the experiments where rodents were to be anaesthetised prior to the use of another method to kill them, CO_2 was chosen. Although CO_2 has anaesthetic properties, its use causes fear and dyspnoea and, depending on the CO_2 fill rate of the killing chamber, it can cause pain. This is why CO_2 should not be used, either as an anaesthetic or as a killing agent, as is further discussed below.

No proposal explained why a certain method was given preference over another one with regard to animal welfare considerations. When investigators had planned a final surgical procedure for sample collection or other terminal experiments, several stated that, if the animal did not reach the planned experimental endpoint and had to be killed prematurely

Figure 2: The methods and agents used to kill conscious animals in a one-step procedure



= *mouse;* = *rat.*

Following 41% of procedures (264/648) where animals were planned to be killed at the end of the experiment, a one-step procedure was proposed.

because a *humane* endpoint had been reached, then another killing method would be used. In many of such cases, CO_2 would be used, without prior administration of a standard anaesthetic agent, as had originally been planned as part of the final surgical procedure.

Even though less humane methods of killing may be authorised, if necessary to achieve the research objectives, the choice of such methods needs to be justified in the proposal, and its impact on the animal taken into account in the project's HBA. However, in our study sample, the choice of method was not accompanied by any scientific explanation or justification. This might be explained by a lack of knowledge about the welfare implications of the methods frequently chosen, such as CO_2.

In the following sections, we cover killing methods that are often proposed, but that should no longer be used in conscious animals. First, we briefly discuss the use of physical methods (decapitation and cervical dislocation) in conscious mice and rats, and the use of chloral hydrate, as well as precautions that should be taken when administering an overdose of the injectable barbiturate, pentobarbital. Due to the extensive research on refinements to the use of CO_2, and the strong likelihood that usage rates are still very high, we reviewed the literature on CO_2 use in detail. Even though further research confirming and strengthening earlier findings on the effects of CO_2 in rodents has been published since the proposals in our sample were submitted, at the time (i.e. in 2010), there was already sufficient evidence supporting the view that CO_2 should not be considered to be a humane killing method (see, for example, the AHWA report from 2005 [50] and the Newcastle Consensus Meeting report from 2006 [51]). Based on that evidence, an official document on the welfare concerns associated with CO_2 use was released by the competent authority of Berlin in early 2009 (52). This document was sent to every animal researcher who applied for a project licence, or notified the competent authority of planning to kill animals for scientific purposes in this federal state.

Decapitation

The debate on how to interpret the electrical activity that persists in the brain for up to 30 seconds after decapitation seems to be no longer an issue, as studies concluded that this activity was not an indication of pain, and that the animals quickly lose consciousness (27). However, due to the distress and fear that rodents might experience when handled, restrained and positioned in the guillotine (27), decapitation should generally only be performed in unconscious animals.

Cervical dislocation

Cervical dislocation (CD) is a common physical killing method for mice and small rats whereby pressure is applied to the neck by pulling quickly to separate the spinal column from the brain. In 2010, cervical dislocation was mostly still regarded as an exception to the general rule of anaesthetising animals prior to killing, because it was assumed that it ensured a quick death and thus, would be less distressing for the animal than undergoing an anaesthesia induction beforehand. However, studies suggesting that cervical dislocation might have a high failure rate had already been published at that time (53, 54). In 2012, Carbone and colleagues (55) verified these concerns. They assessed common methods of cervical dislocation with regard to the time taken to respiratory arrest in anaesthetised mice, and found that over 20% of mice continued to breathe, either beyond the experimental endpoint of 180 seconds (65%) or between 27 and 150 seconds (35%). The authors concluded that cervical dislocation of mice is prone to be unsuccessful in killing the animal, and they highlighted the importance of training in performing this killing method. Importantly, such training should take place on anaesthetised mice only (55). In Switzerland, since 1993, the competent authority had been requiring that, immediately after CD, death of the animal is ensured by exsanguination (56). The Swiss competent authority has recently updated its guidelines and now generally stipulates anaesthesia prior to CD, followed by exsanguination (57). In cases when prior anaesthesia is not feasible, the negative effects for the animals have to be taken into account in the HBA.

Chloral hydrate

Chloral hydrate is not a humane and thus appropriate agent, neither for anaesthesia (see also Herrmann and Flecknell [20]) nor for killing animals. Hence, its use has been rated as non-acceptable by the American Veterinary Medical Association (AVMA) guidelines (27), at least since the 1990s. When used intraperitoneally, as it is the case in rodents, it causes irritation and pain, often resulting in peritonitis, ileus and gastric ulcers. Besides being a strong irritant, animals are likely to experience muscle spasms and gasping and may vocalise as death is caused by a gradual depression of the respiratory system (27).

Pentobarbital

Injection with pentobarbital has been a widely used and accepted method of killing. However, some precautions have to be taken, to ensure rapid and painless death. First, this barbiturate should be administered intravenously whenever possible. When this is not feasible due to the size of the animal, intraperitoneal administration is the norm. Pentobarbital has a pH of 11, and i.p. administration of high pH formulations causes pain. Mixing with lidocaine avoids this problem and represents a potential refinement (58). A sufficient overdose needs to be given, to ensure quick death. A recent study (59) on the humaneness of various agents in use for euthanasia found variable effects after i.p. administration of an accepted overdose (200mg/kg i.p.) of pentobarbital in rats. It was unclear whether this outcome was due to misinjection. Failure rates when injecting intraperitoneally have been shown to be high (6–20% for skilled personnel; 60–63). Chisholm and Pang (59) concluded that the appropriateness of pentobarbital, when administered intraperitoneally, needs further investigation.

Carbon dioxide

Good veterinary practice recommends that animals are anaesthetised before being killed, which was the case for about 60% of the proposed experiments. However, over half of these killings with prior anaesthesia planned to use CO_2 as the anaesthetic agent, and its sole use as a dual anaesthetic/killing agent was the most frequent choice in our sample (Figure 2).

The humaneness of employing CO_2 to both stun and kill animals has been questioned since the 1980s (64). Consequently, it has become one of the most researched areas of *refinement*, with multiple studies demonstrating that its use compromises animal welfare. The ever-growing body of evidence has been showing that rodents find exposure to CO_2 highly aversive (e.g. 59, 65–84). Dyspnoea (68, 69), and fear and anxiety (69, 75, 78, 81–84) are provoked, starting at concentrations from approximately 10% CO_2. The inhalation of concentrations of CO_2 from approximately 40%, at which the animal may still be conscious (depending on the CO_2 fill rate of the killing chamber), causes acidosis, irritation of the mucous membranes (68, 76, 77) and consequently pain, as a result of the formation of carbonic acid on the mucous membranes (59, 71, 79). Furthermore, it has been found that pulmonary haemorrhage and oedema can occur prior to loss of consciousness (78–80).

Fluorinated hydrocarbons, such as isoflurane and sevoflurane, have been recommended as more humane alternatives to induce unconsciousness before the use of CO_2. Critics of the replacement of CO_2 by these alternatives have argued that they are also aversive, and thus their use would not represent an improvement. While fluorinated hydrocarbon anaesthetics become more aversive upon re-exposure, the available evidence suggests that they are

significantly less aversive — indeed, in most cases non-aversive — in naïve animals (70, 85, 86). Physiological evidence in mice, hamsters and the majority of rats suggested that CO_2 exposure induces a significant rise in the plasma levels of the stress hormones epinephrine and norepinephrine, but that plasma levels of these stress hormones were significantly lower when animals were exposed to isoflurane and sevoflurane (71). Additionally, Chisholm *et al*. (87) found that rats exposed to CO_2, but not to isoflurane, showed increased vocalisation, likely due not only to distress, but also due to pain.

There has been uncertainty about the indicator that should be used to ensure loss of consciousness. Moody and colleagues demonstrated that recumbency and absence of movement do not indicate the plane of anaesthesia necessary to not feel noxious stimuli (88). A study that used loss of movement as a proxy for loss of consciousness, found that a return of movement was observed in animals exposed to 100% CO_2 (89). Thus, loss of movement is not a reliable indicator to evaluate the appropriateness of this killing method, as pain or distress experienced by the seemingly unconscious animal can be missed (59). The loss of righting reflex (LORR) appears to be a more reliable indicator for loss of consciousness to a level where noxious stimuli are not felt (59, 90). However, another study has suggested that the researcher should wait about 40 seconds beyond LORR, before switching to 100% CO_2 (88). Bradycardia in rats was identified as sign of irritation and pain (76, 91, 92). A recent study found that bradycardia was present in rats prior to LORR, indicating that the animals did experience pain prior to loss of consciousness (59).

Considering the large and ever-growing body of evidence on the effects of CO_2, it should have already been excluded from use in conscious animals by the time the proposals were submitted in 2010. In 2010, the Canadian Council on Animal Care rated the use of CO_2 as conditionally acceptable, and recommended the use of inhalant anaesthetics prior to killing with this agent (25). However, this is still not properly reflected in the relevant legislation. During the drafting process of *Directive 2010/63/EU*, back in 2005, the expert panel on Animal Health and Welfare (AHAW) of the European Food Safety Authority (EFSA) discussed CO_2 use extensively and concluded that it was not to be used as a sole agent to kill rodents (50). Nonetheless, the expert panel's evidence-based opinion was not fully taken into consideration in the new Directive (19).

Despite the evidence that CO_2 use is inhumane, it is still likely to be the most widely used agent to kill laboratory rodents. A possible explanation is that CO_2 is inexpensive, poses minimal risks to personnel, can be used to kill animals in groups, and does not involve the manual act of killing, which can be distressing for the staff. However, since welfare concerns should be given priority, we submit that competent authorities rigorously scrutinise the scientific justification of continued CO_2 use and permit it only when a plausible scientific justification is given.

Conclusions

In Germany, and at least since the transposition of *Directive 2010/63/EU* also elsewhere in the EU, it is a legal requirement that all *refinement* methods to be used in a study are included in the research proposals. In this retrospective review of research proposals, we detected an overall lack of planned use of *refinement* methods. We propose that this is a strong indicator of the lack of actual use of *refinement* methods in practice, rather than their use being under-reported in animal-based research publications (as is often suggested). Furthermore, this is likely to be a large-scale issue, not just restricted to Germany. We assume that the omission of refinements from the proposals represents either a lack of knowledge of what can be achievably implemented, or a lack of understanding of the importance of refinements to both animal welfare and the quality of the scientific data obtained.

To the best of our knowledge, our study is the first to review the *refinement* methods mentioned in research proposals. Thus, it would be interesting to perform a similar study to assess whether the situation has changed since 2010. The study could also usefully be expanded to include other countries.

The European *Directive 2010/63/EU* now stipulates retrospective assessments of certain animal studies (19), which should help to identify deficiencies and omissions in the application of the Three Rs. However, this task has been allotted to the competent authorities, which were already under-resourced before this new responsibility was added (July 2016, personal communication with Dr Heidemarie Ratsch, head of the competent authority in Berlin from 1996–2016). Hence, a feasible solution could be to outsource these crucial reviews to independent committees comprised of experts, not only in the application of the Three Rs, but also in their respective field of research. Publication of the outcomes of such reviews would enable lessons to be learned and more effective implementation of the Three Rs in the future.

Acknowledgements

We are very grateful to Thorsten Busse for his help in developing an electronic database for comparing the data. We want to acknowledge the German foundation SET, Stiftung zur Förderung der Erforschung von Ersatz- und Ergänzungsmethoden zur Einschränkung von Tierversuchen, for financially support-

ing this project. We also thank ZEBET whose extra travel grant (FK 1329-472) enabled the first author to present the project results at additional conferences. There are no competing interests, and no ethical approval was required.

References

1. Russell, W.M.S. & Burch, R.L. (1959). *The Principles of Humane Experimental Technique*, 238pp. Potters Bar, Hertfordshire, UK: Universities Federation for Animal Welfare. Available at: http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc (Accessed 29.08.18).
2. Buchanan-Smith, H.M., Rennie, A.E., Vitale, A., Pollo, S., Prescott, M.J. & Morton, D.B. (2005). Harmonising the definition of refinement. *Animal Welfare* **14**, 379–384.
3. Grimm, D. (2018). The happiness project. *Science, New York* **359**, 624–627.
4. Richardson, C.A. & Flecknell, P.A. (2005). Anaesthesia and postoperative analgesia following experimental surgery in laboratory rodents: Are we making progress? *ATLA* **33**, 119–127.
5. Coulter, C., Flecknell, P.A. & Richardson, C. (2009). Reported analgesic administration to rabbits, pigs, sheep, dogs and non-human primates undergoing experimental surgical procedures. *Laboratory Animals* **43**, 232–238.
6. Stokes, E., Flecknell, P.A. & Richardson, C. (2009). Reported analgesic and anaesthetic administration to rodents undergoing experimental surgical procedures. *Laboratory Animals* **43**, 149–154.
7. Bara, M. & Joffe, A.R. (2014). The ethical dimension in published animal research in critical care: The public face of science. *Critical Care* **18**(1), R15.
8. Uhlig, C., Krause, H., Koch, T., de Abreu, M.G. & Spieth, P.M. (2015). Anesthesia and monitoring in small laboratory mammals used in anesthesiology, respiratory and critical care research: A systematic review on the current reporting in top-10 impact factor ranked journals. *PLoS One* **10**, e0134205.
9. Carbone, L. & Austin, J. (2016). Pain and laboratory animals: Publication practices for better data reproducibility and better animal welfare. *PLoS One* **11**, e0155001.
10. Bertrand, H.G.M.J., Sandersen, C. & Flecknell, P.A. (2018). Reported analgesic and anaesthetic administration to non-human primates undergoing experimental surgical procedure: 2010–2015. *Journal of Medical Primatology* **47**, 217–225.
11. Pound, P. & Nicol, C.J. (2018). Retrospective harm benefit analysis of pre-clinical animal research for six treatment interventions. *PLoS One* **13**, e0193758.
12. Hooijmans, C.R., Leenaars, M. & Ritskes-Hoitinga, M. (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. *ATLA* **38**, 167–182.
13. Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M. & Altman, D.G. (2010). Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. *PLoS Biology* **8**, e1000412.
14. Würbel, H. (2007). Publications should include an animal-welfare section. *Nature, London* **446**, 257.
15. Anon. (1986). *Deutsches Tierschutzgesetz* in der Fassung vom 1986. [*German Animal Welfare Act* as amended in 1986.] Available at: http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl186s1319.pdf (Accessed 10.12.18)
16. Balls, M. (1999). The biomedical sciences and the need for less inhumane animal procedures. In *Humane Endpoints in Animal Experiments for Biomedical Research* (ed. C.F.M. Hendriksen & D.B. Morton), pp. 1–4. London, UK: Royal Society of Medicine Press.
17. Hendriksen, C., Morton, D. & Cussler, K. (2011). Use of humane end points to minimise suffering. In *The Cost Manual of Laboratory Animal Care and Use* (ed. B. Howard, T. Nevalainen & G. Perretta), pp. 333–353. Boca Raton, FL, USA: CRC Press.
18. Anon. (1972). Deutsches Tierschutzgesetz vom 24 Juli 1972. [German Animal Welfare Act as amended on 24 July 1972.] 29.07.72. Available at: http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl172s1277.pdf (Accessed 10.12.18).
19. Anon. (2010). *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. *Official Journal of the European Union* **L276**, 20.10.2010, 33–79.
20. Herrmann, K. & Flecknell, P.A. (2018). Retrospective review of anaesthetic and analgesic regimens used in animal research proposals. *ALTEX preprint* [doi:10.14573/altex.1804011].
21. Herrmann, K. & Flecknell, P. (2018). Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals: A retrospective review. *ATLA* **46**, 237–289.
22. European Commission (2013). *Report from the Commission to the Council and the European Parliament. Seventh Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union, COM(2013) 859*, 14pp. Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2013:0859:FIN:EN:PDF (Accessed 15.10.18).
23. Metschke, R. & Wellbrock, R. (2002). *Datenschutz in Wissenschaft und Forschung* [Data Protection in Science and Research], 3rd revised edition, 76pp. Berlin, Germany: Berliner Beauftragter für Datenschutz und Informationsfreiheit.
24. Anon. (2006). *Tierschutzgesetz* in der Fassung der Bekanntmachung vom 18. Mai 2006 (BGBl. I S. 1206, 1313), das zuletzt durch Artikel 20 des Gesetzes vom 9. Dezember 2010 (BGBl. I S. 1934) geandert worden ist. [German Animal Welfare Act *as amended on 18 May 2006, last amended on 9 December 2010.]*
25. Charbonneau, R., Niel, L., Olfert, E., von Keyserlingk, M. & Griffin, G. (2010). *CCAC Guidelines on: Euthanasia of Animals Used in Science*, 32pp. Ottawa, ON, Canada: Canadian Council on Animal Care. Available at: https://www.ccac.ca/Documents/Standards/Guidelines/Euthanasia.pdf (Accessed 29.08.18).
26. National Health and Medical Research Council (Australia) and Australian Research Council (2013). *Australian Code for the Care and Use of Animals for Scientific Purposes*. Canberra, Australia: National Health and Medical Research Council. Available at:

https://nhmrc.gov.au/about-us/publications/australian-code-care-and-use-animals-scientific-purposes (Accessed 01.11.18).
27. Leary, S., Underwood, W., Anthony, R., Cartner, S., Corey, D., Grandin, T., Greenacre, C., Gwaltney-Brant, S., McCrakin, M.A., Meyer, R. & Miller, D. (2013). *AVMA Guidelines for the Euthanasia of Animals,* 2013 edition, 102pp. Schaumburg, IL, USA: AVMA. Available at: https://www.avma.org/KB/Policies/Documents/euthanasia.pdf (Accessed 25. 08.18).
28. Fitschen-Oestern, S., Lippross, S., Klueter, T., Weuster, M., Varoga, D., Tohidnezhad, M., Pufe, T., Rose-John, S., Andruszkow, H., Hildebrand, F. & Steubesand, N. (2017). A new multiple trauma model of the mouse. *BMC Musculoskeletal Disorders* **18**, 468.
29. Liles, J.H. & Flecknell, P.A. (1993). The influence of buprenorphine or bupivacaine on the post-operative effects of laparotomy and bile-duct ligation in rats. *Laboratory Animals* **27**, 374–380.
30. Graf, R., Cinelli, P. & Arras, M. (2016). Morbidity scoring after abdominal surgery. *Laboratory Animals* **50**, 453–458.
31. Sauer, M., Fleischmann, T., Lipiski, M., Arras, M. & Jirkof, P. (2016). Buprenorphine via drinking water and combined oral-injection protocols for pain relief in mice. *Applied Animal Behaviour Science* **185**, 103–112.
32. Jirkof, P., Arras, M. & Cesarovic, N. (2018). Tramadol: Paracetamol in drinking water for treatment of post-surgical pain in laboratory mice. *Applied Animal Behaviour Science* **198**, 95–100.
33. Percie du Sert, N., Alfieri, A., Allan, S.M., Carswell, H.V., Deuchar, G.A., Farr, T.D., Flecknell, P.A., Gallagher, L., Gibson, C.L., Haley, M.J. & Macleod, M.R. (2017). The IMPROVE guidelines (Ischaemia Models: Procedural Refinements Of *in Vivo* Experiments). *Journal of Cerebral Blood Flow & Metabolism* **37**, 3488–3517.
34. Wittel, U.A., Wiech, T., Chakraborty, S., Boss, B., Lauch, R., Batra, S.K. & Hopt, U.T. (2008). Taurocholate-induced pancreatitis: A model of severe necrotizing pancreatitis in mice. *Pancreas* **36**, e9–e21.
35. Winston, J.H., Toma, H., Shenoy, M., He, Z.J., Zou, L., Xiao, S.Y., Micci, M.A. & Pasricha, P.J. (2003). Acute pancreatitis results in referred mechanical hypersensitivity and neuropeptide up-regulation that can be suppressed by the protein kinase inhibitor k252a. *Journal of Pain* **4**, 329–337.
36. Hefti, F.F., Rosenthal, A., Walicke, P.A., Wyatt, S., Vergara, G., Shelton, D.L. & Davies, A.M. (2006). Novel class of pain drugs based on antagonism of NGF. *Trends in Pharmacological Sciences* **27**, 85–91.
37. Toth, L.A. (2000). Defining the moribund condition as an experimental endpoint for animal research. *ILAR Journal* **41**, 72–79.
38. Lilley, E., Armstrong, R., Clark, N., Gray, P., Hawkins, P., Mason, K., López-Salesansky, N., Stark, A.K., Jackson, S.K., Thiemermann, C. & Nandi, M. (2015). Refinement of animal models of sepsis and septic shock. *Shock* **43**, 304–316.
39. Peterson, N.C., Nunamaker, E.A. & Turner, P.V. (2017). To treat or not to treat: The effects of pain on experimental parameters. *Comparative Medicine* **67**, 469–482.
40. Hörl, W.H. (2010). Nonsteroidal anti-inflammatory drugs and the kidney. *Pharmaceuticals* **3**, 2291–2321.
41. Ullman-Culleré, M.H. & Foltz, C.J. (1999). Body condition scoring: A rapid and accurate method for assessing health status in mice. *Comparative Medicine* **49**, 319–323.
42. Hendriksen, C.F.M, Morton, D.B. & Cussler, K. (2011). Use of humane end points to minimise suffering. In *The Cost Manual of Laboratory Animal Care and Use* (ed. B. Howard, T. Nevalainen & G. Perretta), pp. 333–353. Boca Raton, FL, USA: CRC Press.
43. Hendriksen, C.F.M., Steen, B., Cussler, K. & Morton, D.B. (1999). *Humane Endpoints in Animal Experiments for Biomedical Research*, Special Edition. Horsham, UK: Laboratory Animals Ltd. Available at: http://www.lal.org.uk/publications/this-is-a-test-subject/humane-endpoints/ (Accessed 30.08.18).
44. Nemzek, J.A., Xiao, H.Y., Minard, A.E., Bolgos, G.L. & Remick, D.G. (2004). Humane endpoints in shock research. *Shock* **21**, 17–25.
45. Hendriksen, C.F.M., Steen, B., Cussler, K. & Morton, D.B. (1999). Preface. *Humane Endpoints in Animal Experiments for Biomedical Research*, Special Edition. V–VI. Horsham, UK: Laboratory Animals Ltd. Available at: http://www.lal.org.uk/uploads/editor/HEP_PREFACE-3.pdf (Accessed 30.08.18).
46. Hawkins, P., Armstrong, R., Boden, T., Garside, P., Knight, K., Lilley, E., Seed, M., Wilkinson, M. & Williams, R. (2015). Applying refinement to the use of mice and rats in rheumatoid arthritis research. *Inflammopharmacology* **23**, 131–150.
47. Wolfensohn, S., Hawkins, P., Lilley, E., Anthony, D., Chambers, C., Lane, S., Lawton, M., Voipio, H. & Woodhall, G. (2013). Reducing suffering in Experimental Autoimmune Encephalomyelitis (EAE). *Journal of Pharmacological & Toxicological Methods* **67**, 169–176.
48. Lidster, K., Jefferys, J.G., Blümcke, I., Crunelli, V., Flecknell, P.A., Frenguelli, B.G., Gray, W.P., Kaminski, R., Pitkänen, A., Ragan, I. & Shah, M. (2016). Opportunities for improving animal welfare in rodent models of epilepsy and seizures. *Journal of Neuroscience Methods* **260**, 2–25.
49 Wolfensohn, S., Hawkins, P., Lilley, E., Anthony, D., Chambers, C., Lane, S., Lawton, M., Robinson, S., Voipio, H. & Woodhall, G. (2013). Reducing suffering in animal models and procedures involving seizures, convulsions and epilepsy. *Journal of Pharmacological & Toxicological Methods* **67**, 9–15.
50. EFSA AHAW Panel (2005). Opinion of the Scientific Panel on Animal Health and Welfare on a request from the Commission related to the aspects of the biology and welfare of animals used for experimental and other scientific purposes. *EFSA Journal* **3**, 292.
51. Hawkins, P., Playle, L., Golledge, H., Leach, M., Banzett, R., Coenen, A., Cooper, J., Danneman, P., Flecknell, P., Kirkden, R., Niel, L. & Raj, M. (2006). Newcastle consensus meeting on carbon dioxide euthanasia of laboratory animals. *Animal Technology & Welfare* **5**, 1–17. Available at: https://www.nc3rs.org.uk/sites/default/files/documents/Events/First%20Newcastle%20consensus%20meeting%20report.pdf (Accessed 30.08.18).
52. Landesamt für Gesundheit und Soziales (2009). *CO_2-Tötung aus Sicht des Tierschutzes*, 1–4.

53. Holson, R.R. (1992). Euthanasia by decapitation: Evidence that this technique produces prompt, painless unconsciousness in laboratory rodents. *Neurotoxicology & Teratology* **14**, 253–257.
54. Keller, G.L. (1982). Physical euthanasia methods. *Lab Animal (NY)* **11**, 20–26.
55. Carbone, L., Carbone, E.T., Yi, E.M., Bauer, D.B., Lindstrom, K.A., Parker, J.M., Austin, J.A., Seo, Y., Gandhi, A.D. & Wilkerson, J.D. (2012). Assessing cervical dislocation as a humane euthanasia method in mice. *Journal of the American Association for Laboratory Animal Science* **51**, 352–356.
56. Bundesamt fur Veterinärwesen (1993). *Richtlinien über das fachgerechte und tierschutzkonforme Töten von Versuchstieren. Richtlinie Tierschutz* 3.01, 17pp.
57. Bundesamt für Lebensmittelsicherheit und Veterinärwesen (2018). *Fachinformation Tierversuche. Fachgerechtes und tierschutzkonformes Töten von Versuchstieren.* 3.01, 14pp.
58. Khoo, S.Y., Lay, B.P., Joya, J. & McNally, G.P. (2018). Local anaesthetic refinement of pentobarbital euthanasia reduces abdominal writhing without affecting immunohistochemical endpoints in rats. *Laboratory Animals* **52**, 152–162.
59. Chisholm, J.M. & Pang, D.S.J. (2016). Assessment of carbon dioxide, carbon dioxide/oxygen, isoflurane and pentobarbital killing methods in adult female Sprague-Dawley rats. *PLoS One* **11**, e0162639.
60. Lewis, R.E., Kunz, A.L. & Bell, R.E. (1966). Error of intraperitoneal injections in rats. *Laboratory Animal Care* **16**, 505–509.
61. Coria-Avila, G.A., Gavrila, A.M., Menard, S., Ismail, N. & Pfaus, J.G. (2007). Cecum location in rats and the implications for intraperitoneal injections. *Lab Animals* (NY) **36**, 25–30.
62. Gaines Das, R. & North, D. (2007). Implications of experimental technique for analysis and interpretation of data from animal experiments: Outliers and increased variability resulting from failure of intraperitoneal injection procedures. *Laboratory Animals* **41**, 312–320.
63. Ballard, T. (2009). Intraperitoneal route of administration — How accurate is this technique? *Animal Technology & Welfare* **8**, 17–18.
64. Freed, D.L.J. (1983). CO_2 euthanasia. *Nature, London* **304**, 482.
65. Kirkden, R.D., Niel, L., Lee, G., Makowska, I.J., Pfaffinger, M.J. & Weary, D.M. (2008). The validity of using an approach-avoidance test to measure the strength of aversion to carbon dioxide in rats. *Applied Animal Behaviour Science* **114**, 216–234.
66. Leach, M.C., Bowell, V.A., Allan, T.F. & Morton, D.B. (2002). Aversion to gaseous euthanasia agents in rats and mice. *Comparative Medicine* **52**, 249–257.
67. Leach, M.C., Bowell, V.A., Allan, T.F. & Morton, D.B. (2004). Measurement of aversion to determine humane methods of anaesthesia and killing. *Animal Welfare* **13**, S77–S86.
68. Leach, M.C., Bowell, V.A., Allan, T.F. & Morton, D.B. (2002). Degrees of aversion shown by rats and mice to different concentrations of inhalational anaesthetics. *Veterinary Record* **150**, 808–815.
69. Hawkins, P., Prescott, M.J., Carbone, L., Dennison, N., Johnson, C., Makowska, I.J., Marquardt, N., Readman, G., Weary, D.M. & Golledge, H.D. (2016). A good death? Report of the second Newcastle meeting on laboratory animal euthanasia. *Animals* (Basel) **6**, E50. Available at: http://www.mdpi.com/2076-2615/ 6/9/50/htm (Accessed 27.08.18).
70. Makowska, I.J. & Weary, D.M. (2009). Rat aversion to induction with inhalant anaesthetics. *Applied Animal Behaviour Science* **119**, 229–235.
71. Marquardt, N., Feja, M., Hünigen, H., Plendl, J., Menken, L., Fink, H. & Bert, B. (2018). Euthanasia of laboratory mice: Are isoflurane and sevoflurane real alternatives to carbon dioxide? *PLoS One* **13**, e0203793.
72. Niel, L. & Weary, D.M. (2007). Rats avoid exposure to carbon dioxide and argon. *Applied Animal Behaviour Science* **107**, 100–109.
73. Niel, L., Stewart, S.A. & Weary, D.M. (2008). Effect of flow rate on aversion to gradual-fill carbon dioxide exposure in rats. *Applied Animal Behaviour Science* **109**, 77–84.
74. Wong, D., Makowska, I.J. & Weary, D.M. (2013). Rat aversion to isoflurane versus carbon dioxide. *Biology Letters* **9**, 20121000.
75. Ziemann, A.E., Allen, J.E., Dahdaleh, N.S., Drebot, I.I., Coryell, M.W., Wunsch, A.M., Lynch, C.M., Faraci, F.M., Howard, M.A., Welsh, M.J. & Wemmie, J.A. (2009). The amygdala is a chemosensor that detects carbon dioxide and acidosis to elicit fear behavior. *Cell* **139**, 1012–1021.
76. Anton, F., Euchner, I. & Handwerker, H.O. (1992). Psychophysical examination of pain induced by defined CO_2 pulses applied to the nasal mucosa. *Pain* **49**, 53–60.
77. Yavari, P., McCulloch, P.F. & Panneton, W.M. (1996). Trigeminally-mediated alteration of cardiorespiratory rhythms during nasal application of carbon dioxide in the rat. *Journal of the Autonomic Nervous System* **61**, 195–200.
78. Cuccheddu, T., Floris, S., Serra, M., Porceddu, M.L., Sanna, E. & Biggio, G. (1995). Proconflict effect of carbon dioxide inhalation in rats. *Life Sciences* **56**, PL321–PL324.
79. Danneman, P.J., Stein, S. & Walshaw, S.O. (1997). Humane and practical implications of using carbon dioxide mixed with oxygen for anesthesia or euthanasia of rats. *Laboratory Animal Science* **47**, 376–385.
80. Ambrose, N., Wadham, J. & Morton, D. (2000). Refinement of euthanasia. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (ed. M. Balls, A.M. van Zeller & M.E. Halder), pp. 1159–1170. Amsterdam, The Netherlands: Elsevier Science.
81. Johnson, P.L., Federici, L.M., Fitz, S.D., Renger, J.J., Shireman, B., Winrow, C.J., Bonaventure, P. & Shekhar, A. (2015). Orexin 1 and 2 receptor involvement in CO_2-induced panic-associated behavior and autonomic responses. *Depression & Anxiety* **32**, 671–683.
82. Mongeluzi, D.L., Rosellini, R.A., Ley, R., Caldarone, B.J. & Stock, H.S. (2003). The conditioning of dyspneic suffocation fear: Effects of carbon dioxide concentration on behavioral freezing and analgesia. *Behavior Modification* **27**, 620–636.
83. Sanna, E., Cuccheddu, T., Serra, M., Concas, A. & Biggio, G. (1992). Carbon dioxide inhalation reduces the function of GABAA receptors in the rat brain. *European Journal of Pharmacology* **216**, 457–458.
84. Winter, A., Ahlbrand, R., Naik, D. & Sah, R. (2017). Differential behavioral sensitivity to carbon dioxide (CO_2) inhalation in rats. *Neuroscience* **346**, 423–433.
85. Moody, C.M. & Weary, D.M. (2014). Mouse aversion

to isoflurane versus carbon dioxide gas. *Applied Animal Behaviour Science* **158**, 95–101.
86. Bertolus, J.B., Nemeth, G., Makowska, I.J. & Weary, D.M. (2015). Rat aversion to sevoflurane and isoflurane. *Applied Animal Behavioural Science* **164**, 73–80.
87. Chisholm, J., De Rantere, D., Fernandez, N.J., Krajacic, A. & Pang, D.S. (2013). Carbon dioxide, but not isoflurane, elicits ultrasonic vocalizations in female rats. *Laboratory Animals* **47**, 324–327.
88. Moody, C.M., Makowska, I.J. & Weary, D.M. (2015). Testing three measures of mouse insensibility following induction with isoflurane or carbon dioxide gas for a more humane euthanasia. *Applied Animal Behaviour Science* **163**, 183–187.
89. Valentine, H., Williams, W.O. & Maurer, K.J. (2012). Sedation or inhalant anesthesia before euthanasia with CO_2 does not reduce behavioral or physiologic signs of pain and stress in mice. *Journal of the American Association for Laboratory Animal Science* **51**, 50–57.
90. Franks, N.P. (2008). General anaesthesia: From molecular targets to neuronal pathways of sleep and arousal. *Nature Reviews Neuroscience* **9**, 370.
91. Kobayashi, M., Cheng, Z.B. & Nosaka, S. (1999). Inhibition of baroreflex vagal bradycardia by nasal stimulation in rats. *American Journal of Physiology* **276**, H176–H184.
92. Yavari, P., McCulloch, P.F. & Panneton, W.M. (1996). Trigeminally-mediated alteration of cardiorespiratory rhythms during nasal application of carbon dioxide in the rat. *Journal of the Autonomic Nervous System* **61**, 195–200.

2.4 Refinement on the way towards replacement: Are we doing what we can?

Citation: Herrmann, K. (2019). Refinement on the way towards replacement: Are we doing what we can?. In: K. Herrmann and K. Jayne, eds. Animal Experimentation: Working Towards a Paradigm Change, Vol. 22, Leiden: Brill, pp. 3-64. https://doi.org/10.1163/9789004391192_002

CHAPTER 1

Refinement on the Way Towards Replacement: Are We Doing What We Can?

Kathrin Herrmann
Freie Universität Berlin, Department of Veterinary Medicine, Institute of Pharmacology and Toxicology, Berlin, Germany; current address: Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, United States
kherrma1@jhu.edu

1 Introduction

> [R]efinement is never enough, and we should always seek further for reduction and if possible replacement.
>
> RUSSELL and BURCH, 1959, Chapter 4

Russell and Burch introduced the principles of replacement, reduction, and refinement of animal experimentation in 1959 in their groundbreaking book, *The Principles of Humane Experimental Technique,* to eradicate inhumanity towards non-human animals (hereinafter referred to as animals). They utilized the term *inhumanity* to indicate negative mental states experienced by animals used in research and the procedures that cause such mental states. Their goal was to avoid the use of animals wherever possible and to improve significantly the treatment of the animals still deemed indispensable, while improving the quality of scientific and medical research and testing (Russell and Burch, 1959). Since the 1990s, the 3Rs have slowly gained more acceptance within the animal research community. They have been recognized by organizations such as the Council of Europe (1986) and the World Organisation for Animal Health (2018), and they have been implemented in law in several countries, for example in Germany and in the UK (Herrmann, Köpernik and Biedermann, 2009; Zurlo, Rudacille and Goldberg, 1996).

Today, the principles are not only embedded in legislation in the European Union (EU) but around the world (Bayne et al., 2015). In the EU, Directive 2010/63/EU on the protection of animals used for scientific purposes came into effect in 2013, thereby requiring all EU Member States to implement the

 | DOI:10.1163/9789004391192_002

3Rs fully. The EU Directive is more far-reaching compared to other legislation since it promotes a strong shift away from animal experimentation, with its goal being "full replacement of procedures on live animals for scientific and educational purposes as soon as it is scientifically possible" (European Parliament, 2010, Recital 10). Furthermore, the EU Directive mandates that replacement should be the first priority, followed by reduction and then refinement to be implemented if animal use is deemed absolutely unavoidable (European Parliament, 2010, Recital 11). Russell and Burch (1959, Chapter 7) proposed the following hierarchy: "Suppose, for a particular purpose, we cannot use replacing techniques. Suppose it is agreed that we shall be using every device of theory and practice to reduce to a minimum the number of animals we have to employ. It is at this point that refinement starts, and its object is simply to reduce to an absolute minimum the amount of distress imposed on those animals that are still used."

As a result of the incorporation of the 3Rs into legislation, which has mainly been driven by ever-increasing societal concerns (cf. Clemence and Leaman, 2016; European Citizens' Initiative, 2016; Jones, 2017; Pew Research Center, 2015, 2018), it would seem reasonable to expect changes within the research industry, particularly replacement of animals with non-animal models. However, the cumulative effect of any such replacements has not prevented the overall number of animals used from steadily increasing since the 2000s (European Commission, 2013; Taylor et al., 2008; Taylor and Rego, 2016). When looking at the 3Rs and their impact, it seems that refinement, the R of *ultima ratio*, is receiving the most attention by the laboratory animal science community (AALAS, n.d.; FELASA., 2016), especially in basic and applied research where the majority of animals are utilized (in the EU, 65% of animals; cf. Daneshian et al., 2015). A survey conducted with participants of laboratory animal science training courses in four European countries found that refinement was seen as more feasible and more pressing than replacement and reduction of animal use (Franco, Sandøe and Olsson, 2018).

Due to this focus, the chapter starts by exploring the application of several refinement methods in practice, commencing with current housing and husbandry standards and a discussion about the benefits of a "culture of care", followed by assessing important experimental refinements. To further assess the quality of animal-based research, it reviews necessary refinements in planning, conduct, and reporting practices of animal studies. The chapter then moves on to look at feasible ways to reduce and replace animal use by, first discussing tools to appraise animal studies whose application could lead to a significant reduction of animal experiments and thus numbers of animals used. It subsequently reflects on what the scientific community has

been doing to move towards replacement of animals in research, testing, and education. Finally, the chapter concludes with recommendations for steps to be taken to work towards using non-animal, human-relevant approaches to biomedical research and testing aimed to protecting human health.

2 Refinement of Animal Housing and Husbandry

> Husbandry is a factor for contingent inhumanity in all types of experiment.
>
> RUSSELL and BURCH, 1959, Chapter 4

Animals used in research, testing, and education spend their lives in a captive environment that is very different from their natural environment. Refined housing gives animals the opportunity to cope with some of the stressors imposed by life in the laboratory (Mason, 2006). Improving their living conditions by trying to meet some of the animals' basic behavioral needs is called *environmental refinement* or *environmental enrichment* (EE). Krech, Rosenzweig and Bennett (1960) were the first to report biochemical changes in the brains of rats kept in a complex housing environment and augmented with daily exposure to novel items in an open field. They coined the term EE when describing this paradigm (Benefiel, Dong and Greenough, 2005). Environmental Enrichment is defined as "[a]ny modification in the environment of captive animals that seeks to enhance the physical and psychological well-being of the animals by providing stimuli which meet the animals' species-specific needs" (Baumans and van Loo, 2013). It includes complex social and inanimate object stimulation (Rosenzweig, 1966). Its positive behavioral effects were first described in rats by Hebb in 1947, who kept them as companion animals in his home. He observed that the rats living in a more complex, stimulating environment learned better and more quickly (Hebb, 1947). In addition to enhancing cognition, EE also promotes neuronal activation, signaling and plasticity in a number of brain regions (Nithianantharajah and Hannan, 2006). In the beginning, research on EE was conducted primarily to assess changes in behavior and brain development. With the increased concern for animal welfare and the establishment of animal welfare science as a specific discipline, has EE been applied to improve the animals' daily lives.

Aside from being driven by animal welfare and health concern, many EE-related research projects have also assessed the influence of poor housing conditions on research data. Garner (2005), van Praag, Kempermann and Gage (2000), and Würbel (2001, 2007), among others, demonstrated that life in barren cages leads to abnormal brain development and to physiological and behavioral malfunction. Standard non- to little-enriched cages can cause a variety

of abnormal behaviors, such as stereotypies (abnormal repetitive behaviour patterns) (see e.g., Würbel and Stauffacher, 1994, 1996; Würbel, Stauffacher and Holst, 1996) and inactivity while awake, observed for example in rhesus monkeys (Hennessy et al., 2014) and mice. Inactivity appears to be an alternative to stereotypic behavior and indicates a depression-like state (Fureix et al., 2016).

Nonetheless, for a period of time, a number of laboratory animal scientists strongly believed that standardizing the animals' environment—by housing animals in barren cages—was essential to control environmental variables (e.g., Bayne, 2005; Eskola et al., 1999; Gärtner, 1999; Tsai et al., 2002, 2003, 2006). The assumption was that standardization was crucial to minimize both variation in the data and the risk of obtaining conflicting results in replicate studies. Many laboratory animal scientists were concerned that implementing EE would add undesirable variation to their responses to experimental treatments (e.g., Bayne, 2005; Eskola et al., 1999; Gärtner, 1999; Tsai et al., 2002, 2003, 2006). However, eight mouse strains kept under such uniform, standardized conditions, and tested on highly standardized behavioral tests in different laboratories, showed significant laboratory dependent variations (Crabbe, Wahlsten and Dudek, 1999). Since then, studies by Augustsson et al. (2003), van de Weerd et al. (2002), Wolfer et al. (2004), and Würbel (2007) have demonstrated that housing conditions can be enriched without increasing variability in experimental results. Additional experiments using mice confirmed earlier research findings that basic environmental enrichments (shelters and nesting material) can be used without compromising the research data (André et al., 2018). Furthermore, this study showed that data from mice who had access to shelters and nesting material is comparable to previous data collected under barren housing conditions, consistent with earlier findings (see Augustsson et al., 2003). The authors concluded that the influence of enrichment on research outcomes was trivial, and that nesting material and shelters could be used without negative impact on study outcomes or loss of comparability to previous data obtained from animals living in impoverished cages. (André et al., 2018).

In the future, rather than using more animals in new experiments on this topic, a systematic review (SR) could be undertaken to provide an overview of the accessible evidence and new knowledge without further animal use. It would also point out knowledge gaps and assess the quality and validity of the conducted animal studies (for more on SRs of animal experimentation, see e.g., Systematic Review Center for Laboratory Animal Experimentation, SYRCLE, n.d. a).

This so-called standardization fallacy (Würbel, 2000), the belief that homogenization of study populations (using the same strain, age, sex, weight, housing conditions, etc.) is an essential part of good experimental design, appears to be one driver for the irreproducibility of results and for the lack of external validity (Bailoo, Reichlin and Würbel, 2014). External validity is the

extent to which experimental results can be used as a basis for generalizations to other human and non-human animal populations in other environmental conditions (van der Worp et al., 2010). This is why authors, including Richter, Garner and Würbel (2009), Richter et al. (2010), Würbel (2000), and Würbel and Garner (2007), promote systematic environmental heterogenization, which is a "controlled and systematic variation of the properties of any given animal (or animal population) and its environment within a single experiment" (Richter, 2017, p. 344). Voelk et al. (2018) compared 440 single- and multi-laboratory preclincial animal studies that had used the same overall number of animals. They compared effect size estimates and found that the studies conducted in one laboratory only, in most cases did not predict effect size correctly, whereas multi-laboratory studies generated more consistent and accurate results. Within-study standardization was identified as a major cause of poor reproducibility. Thus, Voelk et al. (2018) advocate for multi-laboratory design with no increase of overall number of animals being necessary to enhance reproducibility and, potentially, external validity.

EE combined with systematic heterogenization contributes to improved quality of animal experiments (Richter, Garner and Würbel, 2009; Richter et al., 2010; Würbel, 2000; Würbel and Garner, 2007), whereas failure to provide animals with living conditions that meet their species-specific needs jeopardizes both their welfare and experimental validity (e.g., Bailey, 2018; Balcombe, 2010; Bayne and Würbel, 2014; Garner, 2005; Messmer et al., 2014; Olsson et al., 2003; Poole, 1997; Sherwin, 2004; Würbel, 2001, 2007; Würbel and Garner, 2007).

2.1 *Examples of Environmental Refinement*

An example of an extensively researched refinement method is providing mice with various types and sufficient amounts of nesting material to build nests, creating a microclimate needed for breeding and for preventing cold stress (Gaskill et al., 2009, 2012; Gaskill and Garner, 2014; Hess et al., 2008). The thermoneutral zone of mice lies between 26°C and 34°C (Gordon, 1993); and standard temperatures in animal vivariums range between 20°C and 24°C. During their inactive phase, mice prefer temperatures of 30°C–32°C (Gordon, 2012). A proper nest is, therefore, essential for reducing cold stress, which not only compromises animal well-being but also scientific data (Gaskill et al., 2009; Karp, 2012; Messmer et al., 2014). Gaskill et al. (2013) additionally demonstrate its negative effect on breeding performance. Nest building is a species-specific behavior of mice, the absence of which can be used as an indicator of illness (Gaskill and Pritchett-Corning, 2016). Another example involves gerbils, who have a high motivation to dig, since they naturally build and live in burrows. In standard laboratory conditions, where there is not enough substrate to dig tunnels, gerbils show stereotypic digging behavior (Wiedenmayer, 1997). One

solution, based on research conducted by Waiblinger and König (2004), is a nesting box with an attached tunnel. The artificial burrow system seems to help reduce stereotypic digging behavior.

Jirkhof (2015) found that housing conditions that meet the needs of mice help them recover better and faster from experimental procedures. The influence of environment on diseases, such as cancer, has also been demonstrated; for example, by Cao et al. (2010). In colon cancer and melanoma research, mice living in an enriched environment showed reduced tumor growth and increased remission compared to those living in a non-enriched environment (Cao et al., 2010). Rabbits who received special positive attention from their care givers showed a markedly increased resistance to the development of atherosclerosis compared to rabbits who received no extra attention (Nerem, Levensque and Cornhill, 1980).

2.2 *Discussion on Environmental Refinement*

It has been established that animals in a monotonous environment frequently display abnormal behaviors, such as stereotypies (Garner, 2005; Garner and Mason, 2002; Gross et al., 2012; Howerton, Garner and Mench, 2008; Würbel and Stauffacher, 1994, 1996; Würbel, Stauffacher and Holst, 1996). Furthermore, research has demonstrated the importance of environmental refinement, not only for animal welfare and for decreasing the negative health effects of life in captivity, but for its benefits for research outcomes in terms of their reliability, replicability, and validity (e.g., Abou-Ismail and Mahboub, 2011; Garner, 2005; Weed and Raber, 2005).

Due, at least in part, to enforcement of animal protection laws, housing conditions for laboratory animals have improved over the past decade. In the EU, the Commission Recommendation of 18 June 2007 on guidelines for the accommodation and care of animals used for experimental and other scientific purposes (Commission of the European Communities, 2007)—which was later largely adopted by Directive 2010/63/EU (European Parliament, 2010, Annex III)—helped to enhance the captive environment of laboratory animals. However, exceptions to these minimum requirements may be demanded by researchers for certain experiments. Examples include housing social species, such as rats, pigs, or non-human primates, individually and away from their social groups; or not providing rodents with sufficient nesting material and shelters, to allow easier and quicker monitoring. Yet, in most cases, a solution that considers the animals' well-being and does not further compromise their welfare could probably be found.

Moreover, it should be noted that the term most frequently used when talking about an improved living environment, *environmental enrichment*, can be misleading, since it suggests that the standard cage conditions should be

considered normal or species-typical. However, captive conditions have little in common with the natural habitat of every single species used in research. For example, Lahvis (2017) points out that the floor area in a standard mouse cage is 280,000-times smaller than the animal's natural home range. For rhesus macaques, he calculated it is 7 million-fold smaller. Along with the difference in the size of the animals' habitats, the stimulation provided in laboratories is also different from what animals encounter in their natural environments. Burghardt (1996) argues that it would be more accurate to use the term *controlled deprivation*, since all captive environments deprive animals of some natural stimuli. He points out that these restrictions have various, and oftentimes unpredictable, consequences for the welfare of captive animals (Burghardt, 1996, 1999). In fact, a study by Gross et al. (2012) showed that around 12% of mice who lived in enriched cages which contained nesting material, a shelter and a climbing structure, still revealed stereotypic behavior. Moreover, evidence indicates that when stereotypies are not observed, a potential reason could be that they are only displayed when nobody is watching, e.g., in the nocturnal phase (Wells, 2017); or, since highly stereotypic animals seem to cope better than their identically-treated conspecifics, non-stereotypic animals present an even more abnormal, depression-like state as an alternative to stereotypic behavior (Mason, 2006). It has been shown that sustained, uncontrolled stress can, at least in some mouse strains, foster *learned helplessness* (Cabib, 2006).

2.3 *Challenges in the Implementation of Refined Housing*

The enforcement of animal protection laws has contributed to somewhat improved housing conditions for laboratory animals over the past decade. However, despite the mounting evidence of welfare and scientific problems associated with standardized housing, the implementation of animal husbandry knowledge in laboratories has in the author's experience been a major and elusive challenge.

It is increasingly recognized that experimental animals experience serious and repeated stress and distress, caused by life in the laboratory. Besides being a welfare concern, there are multiple factors that adversely affect the animal's biological systems and thus the data collected from these animals (Bailey, 2018). Examples for stressors and thus potential influences on data, besides the confinement itself, include ultrasonic noises (Baldwin, Primeau and Johnson, 2006; Turner et al., 2005), bedding material and cage cleaning (Burn et al., 2006), handling, blood collection, and orogastric gavage (Balcombe, Barnard and Sandusky, 2004), and the experimenters (Chesler et al., 2002) and their sex (Baldwin, Primeau and Johnson, 2006; Sorge et al., 2014).

Numerous studies have shown that animals living in captive environments are generally abnormal and unhealthy, as such environments change their

behavior as well as immune, nervous, and endocrine functionality. Examples include their altered response to infection (Gurfein et al., 2014), altered immune response (Beura et al., 2016; Messmer et al., 2014), increased rates of obesity, Type II diabetes, high blood pressure, and premature death (Martin et al., 2010), altered brain development (Bennett et al., 1964; Kempermann, Kuhn and Gage, 1997; Lewis et al., 2006; Rosenzweig and Bennett, 1969; Rosenzweig et al., 1962), decreased strength and endurance (During et al., 2015), altered sleep, activity patterns, and blood pressure (Martire et al., 2012), altered growth rates (Serrat, King and Lovejoy, 2008), altered organ development, metabolic, growth, and reproduction rates and behavior (Gordon, 2012), and enhanced tumor growth (Cao et al., 2010; Li et al., 2015). As such, untreated control animals do not represent healthy individuals, since they are metabolically abnormal (Martin et al., 2010). To date, there are only a few studies comparing wild versus confined animals, but they all show immense biological differences in physiology, such as structure variation of the visual cortex among caged and free-roaming Norway rats (Campi et al., 2011), lower levels of cholesterol in wild versus captive animals (Schmidt et al., 2006), and immune system dissimilarities (Beura et al., 2016).

We must acknowledge that even if laboratory animal housing is enriched, it cannot be enriched to an extent that it has no negative effect on the animal's welfare (e.g., Burghardt, 1996; Gross et al., 2012). Well-being can only be achieved if the animal experiences positive welfare states, which require a responsive environment the animal can engage with. Studies show that animals prefer complex environments and are motivated to work for them (Anselme, Robinson and Berridge, 2013; Sherwin et al., 2004). Current minimum legal requirements for animal housing in the European Union, laid out in Directive 2010/63/EU, are still insufficient in meeting all needs of all animals; although they are held to be the most progressive in the world. As shown, problems of confinement are manifold. Animals' lives in captivity are monotonous and, therefore, lead to boredom (Burn, 2017; Meagher and Mason, 2012), learned helplessness and depression (Cabib, 2006; Špinka and Wemelsfelder, 2011), and abnormal behaviors. The effects pose serious welfare concerns and raise concerns about the validity and translatability of data obtained from these unhealthy individuals.

2.4 *Potential Improvements*

In assuming an ethical responsibility to improve the lives of captive animals (Gruen, 2014), the goal of husbandry refinement should be not only to reduce stressors but to promote well-being. It is apparent that current housing conditions do not achieve that. One step towards improving animal housing is to provide cages that allow for more natural behaviors. Makowska and Weary (2016a) investigated the frequency of burrowing, climbing, and standing

upright of rats held in pairs in standard (behaviorally restrictive) laboratory cages in comparison with rats in cages allowing these behaviors (larger cages with lower floors, filled with moist soil, holding five rats per cage) over a period of 13 months. Although climbing bouts decreased with age, standing upright and especially burrowing were still frequent behaviors in older rats. Stretching is a corrective response to stiffness caused by immobility or positional stress (Bertolucci, 2011). Makowska and Weary (2016a) found that standard-housed rats performed 9 times more lateral stretches than rats housed in the semi-naturalistic environment. The authors proposed that standard-housed rats were stretching frequently in an attempt to alleviate stiffness from low mobility associated with standard housing. Improved welfare of the rats housed in the semi-naturalistic cages was observed in an anticipatory behavior test that assessed differences in reward sensitivity performed when the rats were 19 and 21 months old (Makowska and Weary, 2016b).

From the animals' perspective, an even better approach would be the radical solution for housing refinement proposed by Lahvis (2017). Lahvis suggests that research animals should live in the wild or at least roam freely in a large, captive environment under naturalistic conditions. He is confident that with available technologies (e.g., cameras, transponders, magnetometers, pressure sensors, global positioning systems), this novel approach could be accomplished for many experiments. Lahvis (2017) advises that biomedical researchers should work together with behavioral ecologists to develop sufficiently complex environments in order to ensure that test subjects produce scientific data not influenced by husbandry.

3 A "Culture of Care" for Animals as Refinement

The term *culture of care* has frequently been referred to by members of the laboratory animal science community to demonstrate "a commitment to improving animal welfare, scientific quality, care of the staff and transparency for the stakeholders." (Norecopa, 2016a). For instance, a working document on the development of a common education and training framework to fulfill Directive 2010/63/EU requirements mentions the culture of care numerous times (National Competent Authorities for the implementation of Directive 2010/63/EU on the protection of animals used for scientific purposes, 2014). Individuals responsible for the welfare of animals should establish and maintain high standards to champion a culture of care among both husbandry and scientific staff (European Commission, 2014). Entire sessions at conferences have been dedicated to this topic, including sessions at the

European Society for Alternatives to Animal Testing (EUSAAT) Congress in 2015 (EUSAAT, 2015) and at the 10th European World Congress on Alternatives and Animal Use in the Life Sciences in Seattle in 2017 (von Aulock, 2017).

Reinhardt (2003, p. 123) identifies compassion for laboratory animals as a refinement: "Kindness and concern for animals in the laboratory often have been stigmatized as subjective, emotional qualities that can undermine the 'objectivity' of biomedical and psychological research." However, since there is evidence that the human-animal bond helps animals to cope with stressful situations in the laboratory (Wolfle, 1987), compassion for laboratory animals should not be dismissed as emotional and subjective but as a sound methodological base for scientifically valid animal-based research (see Mahoney, 1992; Reinhardt, 2003). Compassion implies an acute awareness of an animal's state of emotional, behavioral, and physical well-being and the urge to provide them with the conditions essential for optimal well-being (Reinhardt, 2003). According to Herzog, "there is every reason to believe that individuals who care about their wards on a personal level actually treat the animals better." (2002, p. 30). Morton highlights that, ideally, the staff assessing pain in animals should have an empathetic attitude toward them (Morton, 2000). Such a mindset can also be seen as a protection mechanism to control unrelated, potentially data-influencing, variables (Reinhardt, 2003). Brown (2014) states, "Although there are laws and regulations that govern working with research animals, institutions involved in research, testing, and teaching using laboratory animals should strive to go beyond what is legally required and work to establish a 'culture of care' to ensure animals are treated with compassion and respect." Brown highlights that this culture of care for animals not only benefits animals but the quality of science as well.

3.1 *From Theory to Practice*

How far a culture of care is being implemented on an institutional level is unknown. Personal experiences of this author—as an inspector of animal research institutions in Germany between 2007 and 2016 (Herrmann, 2013; Herrmann and Ratsch, 2010; Maurin, 2012)—revealed differences regarding the level of care for animals within the same institutions, with individual care givers acting more or less compassionately towards their animals. An *institutional* culture of care agenda could not be identified.

The European Commission (EC) (2014) recommends the implementation of such a culture, and other countries have taken steps, in this direction. For example, New Zealand's National Animal Ethics Advisory Committee guide is called *A Culture of Care*: A G*uide for People Working with Animals in Research, Testing, and Teaching* (National Animal Ethics Advisory Committee, 2002). Several pharmaceutical companies, such as Sanofi-Aventis

and Merck (Klein and Bayne, 2007), and commercial breeding companies (Brown, 2014) are reported to have established culture of care programs; however, no external review or assessment of these programs has been published.

3.2 *Towards a Culture of Care and Compassion for Animals*

There is potential for a positive impact of a culture of care on animal use and welfare. But how can we implement such a culture? Schuppli et al. (2017) used a new educational approach to test if exposure to socialized rats, who were trained to fulfill several tasks, fostered compassion among animal experimenters. Six rats were trained using positive reinforcement techniques to, for example, jump onto a scale, or to lift objects. Participants observed these rats and engaged with handling them. After the class, researchers (17) discussed their feelings and reactions. Main findings included that all participants were impressed by the rats' abilities and the close relationship with their trainers. They assumed that this positive animal-human interaction decreased stress in the rats. However, various views existed in regard to potential effects on data. The experimenters expressed unease about emotional difficulties in "sacrificing" their experimental animals after having bonded with them (Schuppli et al., 2017). This highlights one of the major obstacles: When animal researchers develop compassion for their research subjects, they face moral difficulties (see Birke, Arluke and Michael, 2007; Gluck, 2016) and moral harms (see Chapter 13 in this Volume, Johnson and Smajdor, 2019) just as animal caretakers and technicians do. However, this could be an important starting point in moving towards a culture of compassion for all animals which could contribute to their replacement efforts.

4 Refinement of Experimental Procedures

There are several essential refinement methods to reduce the pain, distress, anxiety, and suffering inflicted during the course of experimenting on the animals. Handling and restraint techniques are a source of potential distress and anxiety (Balcombe, Barnard and Sandusky, 2004; Hurst and West, 2010; Meijer et al., 2006); and these techniques have been investigated in experimental studies on stress (Johnson, Sharp and Miller, 2000). To avoid negative effects on behavior, tail handling of mice should be replaced by using tunnels or cupping mice in the open hand (Gouveia and Hurst, 2013). These and other non-aversive handling practices should be implemented industry-wide, since they have been shown to reduce anxiety (Hurst and West, 2010) and optimize the performance of mice in behavioral tests (Gouveia and Hurst, 2017). A recent study by Clarkson et al. (2018) concluded that particular handling methods can

not only cause anxiety, but they can also alter the hedonic value of reward. Tail-handled mice demonstrated a decreased responsiveness to reward and, potentially, a more depressive-like state compared to tunnel handled conspecifics (Clarkson et al., 2018).

For surgical procedures, basic experimental refinements include: proper acclimatization of the animals to the room where anesthesia will be induced (Flecknell, 2018a); optimal anesthesia, peri- and postoperative analgesia; and adequate postoperative monitoring and care, including pain management (Flecknell, 2016; Herrmann and Flecknell, 2018a). The application of humane endpoints also prevents needless suffering. A humane endpoint (or "less-inhumane endpoint," see Balls, 1999, p. 1) represents "[t]he earliest indicator in an animal experiment of (potential) pain and/or distress that, within its scientific context and moral acceptability, can be used to avoid or limit adverse effects by taking actions, such as humane killing, terminating the study, or alleviating the pain and distress." (Hendriksen, Morton and Cussler, 2011, p. 344).

The way an animal is killed is another subject for refinement. Animal care policies in many countries stipulate that death must be painless, and fear and anxiety should be minimized (e.g., Charbonneau et al., 2010; European Parliament, 2010). Less inhumane killing comprises the use of the least distressing and least painful methods that cause rapid loss of consciousness and subsequent death (see e.g., Leary et al., 2013).

The application of our steadily increasing knowledge on experimental refinements should benefit the over 115 million animals who are used annually in research, testing, and education around the world (Knight, 2008; Taylor et al., 2008). However, this benefit cannot be achieved unless the knowledge is translated into practice. In cases where research workers plan to use, for example, less than optimal anesthesia or analgesia protocols, or do not provide other standard veterinary practices, they need to scientifically justify this and demonstrate that the anticipated benefits of the experiments still outweigh the harms inflicted upon the animals (Herrmann and Flecknell, 2018b). Due to the multitude of available means, solutions can be found, in most cases, that help prevent needless animal suffering (Herrmann and Flecknell, 2018a).

4.1 *The Use of Experimental Refinements in Practice*

Several structured and systematic literature reviews have given some insight on certain experimental refinements, notably, killing methods (Pound and Nicol, 2018; Uhlig et al., 2015) and the use of anesthetics and analgesics (Bertrand, Sandersen and Flecknell, 2018; Carbone and Austin, 2016; Coulter, Flecknell and Richardson 2009; Coulter et al., 2011; Pound and Nicol, 2018; Richardson, and Flecknell, 2005; Stokes, Flecknell and Richardson, 2009; Uhlig et al., 2015). For example, animal research involving surgical procedures carried out on

diverse species and published in peer-reviewed journals has been analyzed with regard to analgesic and anesthetic administration (Coulter et al., 2011; Coulter, Flecknell and Richardson, 2009; Richardson and Flecknell, 2005). Stokes, Flecknell and Richardson (2009) focused on studies conducted in two time periods (2000–2001 and 2005–2006), assessing trends in the administration of analgesics and anesthetics to laboratory mice and rats undergoing surgical procedures. The study showed a trend of improvement in terms of safer anesthetic regimens used in the later period examined; however, the findings of this study and an earlier review assessing analgesic use in rodents (Richardson and Flecknell, 2005) show that there was still significant scope for refinement, especially with respect to perioperative care.

A systematic review of anesthesia, analgesia and euthanasia methods used in anesthesiology, respiratory and critical care research in top-10 impact factor ranked journals journals pointed to insufficient reporting of experimental studies with small laboratory mammals. Despite the poor reporting, the review found shortcomings in the application of refinement (Uhlig et al., 2015). Another recent attempt to assess trends in pain management, this time in papers published before 2011 and from 2014 to 2015, further confirmed that reporting (and probably the use) of experimental refinement methods is still poor (Carbone and Austin, 2016). The review demonstrated that scientific publications still cannot be relied upon to present a detailed description of analgesia and anesthesia protocols, not to mention other experimental refinements.

Another approach employed by the author of this chapter, with Flecknell (2018 a, b, c), was to retrospectively review proposals for authorization of basic and applied animal research studies to learn which experimental refinements were proposed. Over 500 applications submitted to the German competent authorities in 2010 were reviewed. German law stipulates that all possible refinements that are planned in an animal study are described in detail in its proposal. The review's goal was to evaluate the intended application of and, thus, the awareness about possible refinements. Among other results, postoperative analgesia was not proposed for 30% of surgeries; and, in the majority of cases, its scientific necessity was not further discussed. Following 10% of procedures, animals were to be given pain relieving medication only if the investigators decided that it was necessary; however, structured assessments to detect pain were absent (Herrmann and Flecknell, 2018a).

4.2 *Discussion on Refinement of Experimental Procedures*

Structured and systematic literature reviews and the work of this author found strong indications for flaws in the administration of experimental refinement. Refinement methods need to be fully employed in order to minimize stressors that can lead to distress, such as suffering from postoperative pain, or living

in a barren cage. The biological consequences of stress and distress compromise rigor, reliability, and relevance of data collected from these animals (see Bailey, 2018 for a review on how stress of laboratory life and experimentation can adversely affect research data). Animal researchers are responsible for the animals they use (National Health and Medical Research Council, 2013). Thus, they and their animal care staff should know enough about animal behavior to properly assess the health and well-being of their test subjects. In the European Union, they are legally required to be skilled, educated, and equipped to detect and relieve suffering accordingly (European Parliament, 2010, Article 24).

There are several challenging areas of refined care and use that should be addressed. For example, there is a need for automated, remote, 24/7 cage-side monitoring to identify abnormal behavior, which is especially important when assessing the welfare of genetically modified animals, as well as for prey species who tend to mask their medical condition or psychological state. Additionally, there is a need for further development and implementation of valid pain-assessment techniques to determine the efficacy of treatment in the individual animal due to individual variations in pain response. While there is necessity for further research into certain areas of experimental refinement, it is essential that we apply the knowledge we already have, so that immediate improvements in animal welfare can be achieved.

5 Refinement of Experimental Design, Conduct, and Reporting

There have been quality problems throughout medical and biomedical research (Begley and Ellis, 2012; Harris, 2017; Pound and Bracken, 2014; Prinz, Schlange and Asadullah, 2011). "The scandal of poor medical research" with human subjects was discussed in a British Medical Journal (BMJ) editorial in 1994 (Altman, 1994). A biostatistician took a prominent stance against the unethical misuse of statistics (Altman, 1980). In a follow up 20 years later, another BMJ editorial called, "Medical research—still a scandal," concluded that matters have become worse (Smith, 2014). It is apparent that the quality of *in vivo* research with animal and human subject demands urgent improvement. Weaknesses in design, conduct, and analysis of biomedical and public health research studies yield misleading results and, thus, waste resources (Ioannidis et al., 2014). Since legally-required animal data forms the basis of decisions to move forward to human clinical trials, flawed animal research is additionally problematic.

Aside from evidence that many animal experiments that are performed never get published (Scherer et al., 2018), a large part of what gets published is

incorrect (e.g., Harris, 2017; Ioannidis, 2005; Freedman, Cockburn and Simcoe, 2015). Ioannidis (2005) argues that it is highly probable that most published findings are indeed false. He drew his conclusion after conducting simulation studies and SRs. He calculated that, at best, only one in three publications took basic precautions to minimize bias (Ioannidis, 2005). Freedman, Cockburn and Simcoe (2015) estimated that more than 50% of all preclinical studies in the United States are unreliable, and that the financial damage of these irreproducible preclinical studies is US$28 billion per year. Their analysis revealed that about 20% of the studies had an untrustworthy experimental design, one quarter used media that contained contaminated cells and antibodies, and in 18% of studies the data analysis was poor. All of these issues have contributed to the so-called *reproducibility crisis* in animal research (e.g., Aarts et al., 2015; Baker, 2016; Begley and Ioannidis, 2015; Bracken, 2009; Collins and Tabak, 2014; Freedman, Cockbury and Simcoe, 2015; Ioannidis, 2005; Perel et al., 2007; Pound et al., 2004; Pound and Bracken, 2014; Reichlin, Vogt and Würbel, 2016; Scannel and Bosley, 2016; Würbel, 2016). A review of the literature by Bailoo, Reichlin and Würbel (2014) strongly suggests that experimental design and conduct of laboratory animal research are in need of improvement. A study by Vogt et al. (2016) revealed that animal researchers working in Switzerland do not apply basic principles of study design to avoid bias and do not properly report their study outcomes. They also found that neither the Swiss regulatory authority nor the international journals and their peer reviewers had adequate knowledge to recognize these flaws.

In an attempt to improve the quality of research reports, several checklists and guidelines have been put in place, such as Consolidated Standards of Reporting Trials (CONSORT) and Standards for Reporting of Diagnostic Accuracy (STARD) for human clinical trials, Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) for SRs and meta-analyses, and Gold Standard Publication Checklist (GSPC), Animal Research: Reporting of *In Vivo* Experiments (ARRIVE) guidelines for animal research (Glasziou et al., 2014; Hooijmans, Leenaars and Ritskes-Hoitinga, 2010) and HARRP, the harmonized animal research reporting principles, which are a recent attempt by ICLAS (International Council for Laboratory Animal Science) in harmonizing animal research reporting to further improvements in the scientific rigor of animal experiments (Osborne et al., 2018). The ARRIVE guidelines (Kilkenny et al., 2010) are the most widely known for reporting of animal-based experiments. These guidelines have recently been complemented by the Planning Research and Experimental Procedures on Animals: Recommendations for Excellence (PREPARE) guidelines, which help to ensure quality when preparing animal studies (Smith et al., 2017).

The ARRIVE guidelines were adopted by more than 1000 scientific journals, and more than 20 funding agencies were expected to endorse them in 2010 (Baker et al., 2014; Enserink, 2017). Two years later, Baker et al. (2014) assessed the degree to which they had been endorsed by reviewing journals, such as Nature and PloS, and found that there was little improvement. The knowledge about and the use of reporting guidelines, such as the ARRIVE guidelines, is still not widespread, as a study by Reichlin, Vogt and Würbel (2016) has shown. Reichlin et al. asked animal experimenters in Switzerland to complete a questionnaire regarding their use of measures against risk of bias. Only 16% responded. The ARRIVE guidelines were known by less than half (43.7%). Furthermore, Carbone and Austin (2016) found no increase in reporting of analgesic use in the articles published in journals that had agreed to endorse the ARRIVE guidelines. Results of a recent randomized controlled trial of close to 1,700 scientists, who submitted papers to the scientific journal *PLoS One*, suggest that scientists are either ignoring the guidelines or are still unaware of their existence. Another finding was that, even when an ARRIVE checklist was completed, the correlating papers were actually not more compliant, which may indicate that researchers do not know what is expected of them and why providing this information is crucial, emphasizing the importance of proper training (Enserink, 2017; Hair et al., 2018).

5.1 *Sources of Bias in Animal-based Research*

There is a large number of potential sources of bias in animal research. Not surprisingly, most published animal studies have some risk of bias (Macleod et al., 2015). Safeguards to avoid bias in study design, conduct, and analysis include randomization of treatment groups to eliminate systematic differences between them, blinding of investigator to treatment and to handling of data, and reporting on sample size estimation (Macleod, 2011). Analytical errors may account for close to a quarter of the irreproducible studies (Freedman, Cockburn and Simcoe, 2015); thus, knowledge about statistical methods is essential. Other suggested items for reporting include: a clear description of the hypotheses tested or primary and secondary objectives of the study, housing and husbandry, including welfare-related assessments and interventions, adverse events, and interpretation of results, taking into account the hypotheses/study objectives (Kilkenny et al., 2010). Part of the reproducibility and translatability crisis is considered to be due to poor experimental design and conduct of animal experiments (Bailoo, Reichlin and Würbel, 2014; Ioannidis, 2005; Macleod, 2011; van der Worp, 2010; Würbel, 2016), including the influences of inappropriate animal housing (Lahvis, 2017) and handling (Gouveia and Hurst, 2017), insufficient pain relief (Carbone and Austin, 2016; Herrmann and Flecknell, 2018a),

as well as the absence of other refinements, such as careful monitoring, early humane endpoints, and less inhumane killing methods to reduce pain, suffering, and distress (Herrmann and Flecknell, 2018b and 2018c). The other, and perhaps larger, part is due to insurmountable species differences (Pound and Bracken, 2014; Pound and Ritskes-Hoitinga, 2018), which Russell and Burch already discussed 60 years ago (1959, Chapter 5).

Another source of bias is selective reporting when publishing results of animal experiments (Briel et al., 2013; Ioannidis, 2012; Landis et al., 2012; Lees et al., 2012; Macleod et al., 2004; Pound and Bracken, 2014; Sena et al., 2010; Tsilidis et al., 2013; Würbel, 2016). One problem relates to *negative* findings—studies for which the original hypotheses were not proven. Some of these are not published at all, which has long been recognized as a source of publication bias. The second problem relates to studies that are reported incompletely. For example, only the parts that demonstrate that the treatment is effective are reported, with whole experimental groups excluded from reporting. This is selective outcome and analysis reporting bias (Ioannidis, 2012). These partially or unreported studies may be repeated by others and thus represent an unnecessary waste of animal lives. Incomplete reporting of published findings makes it impossible to replicate studies (Begley and Ellis, 2012). Because negative findings are often not published (Scherer et al., 2018), the value of published findings is over-estimated, which, in part, could explain some of the difficulties in translating promising preclinical results into effective therapies for human disease (Bath et al., 2009; Mergenthaler and Meisel, 2012; Sena et al., 2010).

Yet another pitfall is researchers' freedom of flexibility in data collection, analysis, and reporting, which dramatically increases false-positive rates in the literature and, therefore, contributes to misleading animal research data and overestimation of its significance. Regardless of the nominal endorsement of a maximum false-positive rate of 5% ($p \leq .05$), standards for disclosing details of data collected and analyzed make false positive results very likely (Simmons, Nelson and Simonsohn, 2011). The authors describe this as *p-hacking*. Oftentimes, an experimenter is more likely to find evidence that an effect exists falsely than to find evidence that it does not correctly. This occurs because of the investigators' degree of freedom with regard to the amount of data collected and analyzed, the exclusion of certain observations made, the comparison or combination of conditions, the variables considered, and so forth. It is uncommon for researchers to make these decisions before undertaking experiments. Their exploratory behavior is explained as ambiguity in how best to make these decisions and the desire to find statistically significant results.

Confirmatory bias is another potential pitfall, since people tend to interpret ambiguous information in such a way that it supports a justifiable

conclusion that matches their own aspirations (e.g., Dawson, Gilovich and Regan, 2002). *HARKing* (i.e., Hypothesizing After the Results are Known) (Kerr, 1998) is another common and problematic practice in science. Statistical tests to differentiate true effects from random noise are designed for confirmatory research, not exploratory research. Thus, when researchers change their *a priori* hypotheses after obtaining their results, this leads to false conclusions.

An additional area that urgently needs refinement is transparency and data sharing to avoid publication bias and needless repetition of studies. Openness is a cornerstone of science and could help in reducing the reproducibility problem science is facing (Errington et al., 2014; Harris, 2017; McNutt, 2014). It is essential to discover and correct errors. The Food and Drug Administration Modernization Act of 1997 (FDAMA) (Food and Drug Administration, 2018) requires scientists to register their hypotheses and endpoints in advance, if they plan to run a clinical trial on potential new pharmaceutical drugs (ClinicalTrials.gov). This new law went into effect in 2000. It also requires pharmaceutical companies to publish their results, thus, avoiding publication bias. Despite the insufficient enforcement of the law, as many scientists still do not report the results of their studies (Harris, 2017), the indispensability of such provisions is demonstrated by the findings of a study conducted by Kaplan and Irvin (2015). They assessed whether the FDAMA had any effect on study outcomes. Before the law was in place, 57% of drugs or supplements showed benefits; after the law was in place, 8% of the studies published confirmed their preregistered hypotheses (Kaplan and Irvin, 2015). Such a prospective registration process is currently exceptional for animal-based studies, but it is unquestionably required in order to enhance transparency, reduce selective reporting bias, and prevent duplication. The Center for Open Science, a nonprofit where researchers can register their hypotheses *a priori* (https://cos.io) and Preclinical Trials, a platform for registration at the outset of all types of animal studies (www.preclinicaltrials.eu), will hopefully improve the current situation. An additional measure to improve transparency, and potentially reproducibility, is data sharing, which is a requirement for publication by some major journals but many researchers still refuse to share. By sharing data, errors can be discovered (e.g., Salzberg et al., 2001). This is especially important in animal research, since it helps reduce the number of animals used and sheds light on the real value of animal derived data.

5.2 *Necessary Steps*

The improved quality of human clinical trials was achieved by strategies to minimize bias, *a priori* power analysis and further biostatistics, clear definition of the primary and secondary endpoints, data monitoring and auditing, internationalization and inclusion of multiple centers, external steering committees and safety monitoring, rigid publication standards, trial registries, and

more (Dirnagl and Fisher, 2012). The lessons learned from the improvement of human clinical trial quality should be adopted by preclinical (Dirnagl and Fisher, 2012) and all other biomedical research fields (Hartung, 2013), where relevant and with appropriate changes, since flawed research is unscientific and unethical. The ethical issues with research involving animals become extra critical as needless animal suffering must be avoided, and as preclinical animal data generally forms the basis for decisions whether to proceed to human clinical trials. Thus, in order to adhere to the 3Rs, the following efforts are crucial:

- Education and ongoing training of researchers in experimental design, statistical methods, and model selection (Justice and Dhillon, 2016).
- Close assistance in study design by institutional animal welfare bodies and by biostatisticians.
- As a possible solution for the problem of false positives making their way into the literature, some researchers suggest the *p*-value threshold should be reduced to 0.005 (Chawla, 2017). Others say researchers should select and justify *p*-value thresholds for their experiments, before collecting any data. These levels should be based on factors such as the potential impact of a discovery. These thresholds could then be evaluated via *registered reports*, a type of scientific article in which methods and proposed analyses are peer-reviewed *before* any experiments are conducted (Chawla, 2017).
- Transparency must be improved as it is crucial to document all anticipated or exploratory steps in the study. Prospective registration of all animal studies with their hypotheses and endpoints is essential to prevent selective-reporting biases (Ioannidis, 2012) and avoid study duplications (Preclinical Trials, n.d.).
- Disclosure and openness are critical elements of science for self-correction, and they can help avoid poor practices, such as HARKing.
- The use of preparation and reporting guidelines, such as the PREPARE guidelines (Smith et al., 2017) combined with the ARRIVE guidelines (Kilkenny et al., 2010), should be a mandatory, legally required part of funding applications, project license applications, as well as publications. Education on how to fill out the checklists and present the required information in the publication, as well as a focus on enforcement of compliance to both by journals, is critical (Eisen, Ganley and MacCallum, 2014; Enserink, 2017; Hair et al., 2018; Herrmann and Flecknell, 2018a).
- Raw data, analyses, and protocols must be made available to allow other researchers to verify results. This can easily be achieved by using data repositories (e.g., https://datadryad.org or https://figshare.com).
- Reporting of all study outcomes to avoid traditional reporting bias and selective outcome and analysis reporting bias should be mandatory (Ioannidis, 2012).

– Retrospective assessments of animal studies (see EC Expert Working Group for Project Evaluation and Retrospective Assessment, 2013, pp. 28–32) should be performed comprehensively and by independent experts; and all results should be published to enhance transparency, minimize publication bias, identify animal models lacking external validity, and, thus, improve future research.
– Mandatory data sharing so that other scientists can build on the work and discover errors faster (cf. the error in the Human Genome Project discovered by Salzberg et al., 2001). Data sharing should be compulsory, especially when research is publicly funded.

It is equally important that funding and regulatory bodies, animal ethics committees, animal welfare bodies, journal editors, and peer reviewers have a detailed knowledge of these topics in order to recognize flawed research studies. This requires effective and thorough education and training of funders, animal ethics and welfare committees, and regulatory body members on how to assess animal research proposals (Vogt et al., 2016). Furthermore, in order to review these applications in-depth, enough time and manpower are a prerequisite.

6 Refinement: Are We Doing What We Can?

As presented in this chapter, knowledge about and implementation of refinement of husbandry, experimental procedures and design, conduct, and reporting appears to still be patchy. Since adoption of refinement strategies has been inconsistent, it would seem that rather than use additional animals to carry out more refinement research, we should focus on the comprehensive application of existing refinements in animal laboratories as well as on reducing and replacing animals.

6.1 *But What about the Refinement of Animal Models?*

Animal models ought to describe a biological phenomenon that the model species has in common with the target species. Significance and validity, in terms of the translatability of results produced in an animal model to the human condition, "depend on the selection of a suitable animal model," writes Hau (2008, p. 4), which is why comprehensive knowledge about comparative anatomy and physiology is essential. A majority of animal models developed with the expectation to study the origin, disposition, and treatment of human disorders and is created through experimental induction, genetic modification, or breeding of disease-causing mutations (Hau, 2008, p. 4). These presumed predictive models are used to find treatments or to assess the toxicity of drugs and other chemicals (Hau, 2008). Hence, they cause conditions

associated with pain and distress up to severe, long-lasting suffering for these animals.

Some laboratory animal scientists focus on the refinement of animal models in an attempt to reduce the suffering caused. Examples for refinement recommendations of animal models include those described for mice and rats who are utilized as models of ischemic stroke (Percie du Sert et al., 2017), for rheumatoid arthritis (Hawkins et al., 2015), in experimental autoimmune encephalomyelitis (EAE) (Wolfensohn et al., 2013a), as models and in procedures involving seizures, convulsions, and epilepsy (Lidster et al., 2016; Wolfensohn et al., 2013b), and as models of sepsis and septic shock (Lilley et al., 2015). If the gathering of such recommendations does not involve additional harmful animal experiments, and in case these guidelines are then applied in practice, they could lead to an improvement of the individual animal's life.

However, due to failure of numerous models to predict human outcomes (e.g., Joffe et al., 2016; Mak, Evaniew and Ghert, 2014; Pharmaceutical Research and Manufacturers of America, 2015, 2016), and due to limited funding, it seems crucial to first assess carefully which research methods and models to use. In the case of sepsis models, for example, there have been multiple publications highlighting the differences in human and mouse immunology (e.g., Mestas and Hughes, 2004; Rittirsch, Hoesel and Ward, 2007; Seok et al., 2013; Shay et al., 2013; Payne and Crooks, 2007). After over 20 years of unsuccessful research in this field, a number of scientists finally investigated why, out of the approximately 150 new compounds that were developed for the treatment of sepsis using mice, not one had beneficial effects for humans. They identified around 5,000 genes that are activated or deactivated by inflammation in humans who suffered from sepsis, trauma, or burns. They went on to look for the same genes in one commonly-used strain of mice and realized that there was no correlation (Seok et al., 2013). As a consequence of the dissimilarity of mouse and human immune systems, the entire field of sepsis research in mice has been called into question, regarding its predictive value for humans. Paradoxically, funding for this kind of animal research, which is also known for causing severe levels of animal suffering, is still ongoing (Leist and Hartung, 2013). At the same time, human-based sepsis research has led to clinical trials of effective therapies (van der Poll, 2012).

7 Reduction and Replacement: Are We Doing What We Can?

Most animal research is being justified as indispensable to furthering human healthcare. However, despite measures being taken to improve the quality of animal-based research, the translational success rate from animal studies to

humans is low: Less than 12% of drugs entering clinical trials result in an approved medicine (Pharmaceutical Research and Manufacturers of America, 2015, 2016); and between 51% and 89% of preclinical studies are not reproducible (Freedman, Cockburn and Simcoe, 2015; Harthorne and Schachner, 2012). There is an ongoing debate among scientists as to why animal models fail to be predictive: Is this mainly due to poor scientific rigor and reporting, to species differences, or to the fact that today we mainly deal with complex, oftentimes, chronic ailments of which many are not well understood and, thus, impossible to model in other animals?

As a consequence of the failure to translate findings to humans, new criteria for mouse models have been described (Justice and Dhillon, 2016). Hoping to enhance animal models of stroke, Dirnagl and Fisher (2012) call for international, multicenter, preclinical Phase III-type studies of promising new ischemic stroke therapies in animals before moving to clinical trial. As Phase III studies would be based on prior studies and would use various strains and species (Dirnagl and Fisher, 2012), as well as older animals with various comorbidities (e.g., diabetes mellitus, obesity, and hypertension) (Mergenthaler and Meisel, 2012), the severity of these experiments and the numbers of animals involved would markedly rise. Of close to 100 interventions that improved the outcome in animal stroke models, which were tested in clinical trials, only one intervention improved the outcome in human patients (O'Collins et al., 2006). Despite decades of research, most translational stroke trials that aim to extrapolate basic research findings into clinical treatments, particularly in the area of neuroprotection, have failed (Mergenthaler and Meisel, 2012). The authors admit that, to date, there is no ideal animal model for stroke, and that more complex models are needed to improve translational success in experimental stroke research (Mergenthaler and Meisel, 2012). Thus, at the time of writing, Mergenthaler and his colleague Stachelscheid are developing human stem cell-derived 2D and 3D models for stroke (vfa, 2017). Building on the latest *in vitro* research to model human brain development and disease, they plan to employ a recently established protocol for generating 3D brain tissue, so-called cerebral organoids, from human pluripotent stem cells that can be applied to study a number of human brain diseases (Lancaster and Knoblich, 2014). Renner et al. (2017) further examined the development and potential differentiation of cerebral organoids, which hold great potential to advance human-relevant stroke research.

7.1 *Potential for Reduction by Critical Appraisal of Animal Studies*

Several unsuccessful animal models have been discussed, such as for Alzheimer disease (Cavanaugh, Pippin and Barnard, 2014; Pippin, Cavanaugh and

Pistollato, 2019, Chapter 20 in this Volume; Pistollato et al., 2016), for stroke (Shuaib et al., 2007; van der Worp et al., 2010), for tuberculosis (Fonseca et al., 2017); for asthma (Mullane and Williams, 2014), for HIV/AIDS, for neurological, menopausal human therapy, and for cancer research as well as drug development (Pippin, 2012). Since only disease models with high predictive validity are likely to yield positive results and treatments for humans, it is critical to assess the reliability, reproducibility, and validity of the animal model first. With the overall low quality and predictive validity of the majority of research studies, it has become evident that animal-based studies require rigorous evaluation (Pound et al., 2004). A solid methodological approach would be to systematically review and to perform meta-analyses of animal studies, as SRs are seen by experts in the field of evidence-based medicine as the highest level of medical evidence (Hooijmans, Leenaars and Ritskes-Hoitinga, 2010).

7.1.1 Systematic Reviews (SRs)

A systematic review (SR) is a literature review that focuses on a specific question with the aim to identify and assess all relevant studies in order to generate new, high-quality evidence. Thus, it enables evidenced-based decision making (Norecopa, 2017). A SR may contain a meta-analysis. In a meta-analysis, the results of a number of independent studies are statistically combined to calculate the average effect of studies addressing the same question, which may lead to more reliable conclusions and may help to minimize needless duplication of animal studies (Hooijmans et al., 2014a). SRs conform with the implementation of the 3Rs concept (Ritskes-Hoitinga, 2016), as their application leads to a more evidence-based choice of animal models (e.g., de Vries et al., 2012; Sloff et al., 2014; Zeeff et al., 2016). They help decrease unnecessary animal studies, the evidence they produce should further responsible animal use, and they increase scientific quality (van Luijk, 2016), as they are an excellent tool to assess study quality by evaluating the internal, external, and construct validity of the models. Internal validity is the degree to which the design, conduct, and analysis of the experiment remove potential bias, so that the interpretation of a causal relationship between an experimental treatment and variation in an outcome measure is secured (Bailoo, Reichlin and Würbel, 2014). The extent to which animal data gives a basis for generalization to other animal and human populations, including other environmental circumstances, represents the external validity; and construct or predictive validity shows how good the model is, the rate to which the sampling properties are representative for the entities they ought to represent (Bailoo, Reichlin and Würbel, 2014; Würbel, 2017). An example for a SR on internal validity is the study of Macleod et al. (2008) and for construct validity, the work of Sena et al. (2010), both focusing on reasons

for translational failure of experimental stroke. SRs are a significant tool to identify quality issues with primary animal studies. For example, a recent SR on the welfare implications of toe clipping and ear notching revealed that the underlying animal experiments were too flawed to draw conclusions (Wever et al., 2017). SRs are excellent to assess the risk of bias in animal studies and thus to evaluate the reliability of the available evidence (van Luijk et al., 2014). Perel et al. (2007) systematically reviewed the success of treatments in animals and in humans, with head injury, hemorrhage, thrombosis due to acute ischemic stroke, acute ischemic stroke, and osteoporosis as well as preventive medication in neonatal respiratory distress syndrome, with their applications in humans with these impairments. Their conclusion was that the incongruity between animal and human studies may be due to bias or to the failure of the animal models to mimic clinical disease (Perel et al., 2007).

SRs of animal studies are still much less common than in the clinical setting, where they are frequently used to make evidence-based decisions on healthcare; but awareness of the benefits of the utility of SRs of animal research has been increasing (Hooijmans et al., 2014; van Luijk et al., 2014). The Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES) group, at the University of Edinburgh in the UK, and the Systematic Review Center for Laboratory Animal Experimentation (SYRCLE), at Radboud University Medical Center in the Netherlands, provide a supporting framework for groups who are or want to get involved in the SR and meta-analysis of data from experimental animal studies and offer advice and training (CAMARADES, 2014; SYRCLE, n.d. b). SYRCLE has published a step-by-step guide on how to identify all relevant animal studies (Norecopa, 2017), as well as a tool similar to the Cochrane tool for assessing risk of bias in randomized clinical trials (Higgins et al., 2011), to assess the risk of bias in animal-based studies (Hooijmans et al., 2014). It is important to receive proper training first, as one needs to be aware of the pitfalls and limitations of these tools, and how they can be misused and/or misleading (Gurevitch et al., 2018). Various types of reporting biases, together with the limited methodological quality of some studies on which meta-analyses and SRs are based, can impede their conduct and interpretation (e.g., Benatar, 2007). When publication bias against negative animal studies exists, it will lead to an overestimate of the value of animal studies. It is likely that if unpublished studies were to be included, then SRs would show more studies with no effect in animals (Akhtar, Pippin and Sandusky, 2009). Checklists and tools have been proposed to help improve SRs and meta-analyses (Hooijmans et al., 2014; Moher et al., 2009).

The use of SRs should be standard practice within animal-based research, in the same way it has become a vital part of clinical research (Hooijmans, Leenaars and Ritskes-Hoitinga, 2010; Hooijmans and Ritskes-Hoitinga, 2013;

Pound and Bracken, 2014; Pound et al., 2004; Sandercock and Roberts, 2002). SRs should be conducted prior to a new animal study to assess the validity of the proposed animal model and to avoid needless animal use (Ritskes-Hoitinga and Wever, 2018). For example, in refinement research, SRs are an efficient way to gather new high-quality data without having to experiment on additional animals. As shown in this chapter, the implementation of new knowledge about refinements to improve animal welfare has proven very difficult. A prominent example is the use of carbon dioxide to kill animals. Extensive research conducted on this welfare topic has produced overwhelming evidence against its use, but these findings still have not led to the abolishment of this common practice. At the time of writing, Turner et al. are conducting a SR on the use of carbon dioxide as a killing method for mice and rats. Their protocol (SYRCLE, n.d. c), as well as the protocols of others, are published on the SYRCLE website and, since 2018, protocols of SRs relevant to human health can be registered at the international prospective register of SRs, called PROSPERO (https://www.crd.york.ac.uk/prospero/).

7.1.2 Other Retrospective Assessments

Conducting retrospective assessments (RAs) is a useful way to identify disease models and research methods that may be of limited value. Since 2013, RAs are mandatory for certain animal studies in the European Union (European Parliament, 2010, Article 39). Members of the animal research inspectorates have been required to assess the outcomes of animal studies that were classified as severe and/or use non-human primates. The animal researcher has to submit the necessary documents so that the competent authority can evaluate whether the study objectives were met, the actual harm inflicted, and whether the severity of procedures coincided with the prospective assessments, and the number of animals used. In addition, the competent authorities must appraise any component that can advance the implementation of the 3Rs (European Parliament, 2010, Article 39).

These RAs could be extremely effective in facilitating a critical review of the use of animals in scientific procedures, if there are sufficient and qualified personnel to conduct them, as the EC's aim with these RAs is to identify 3Rs improvements and enhance transparency to the public (EC Expert Working Group for Project Evaluation and Retrospective Assessment, 2013). Publication of RA results of all studies, including those that produced negative results and may not be published elsewhere, would likely be of significant value. It would increase the knowledge base in a range of disciplines, reduce risks of duplication of studies, and inform the design of future research (EC Expert Working Group for Project Evaluation and Retrospective Assessment, 2013). However, only about one sixth of all EU Member States agreed to make the RA results publicly

available. This is not enough to meet the EC's goals. To achieve maximum benefit, access to study results should be given not only to regulatory authorities but to independent experts, in order for them to perform critical reviews of these data. And all RA results have to be made publicly available. It is possible to do so and still protect intellectual property by redacting and anonymizing certain parts of the documentation.

7.1.3 Necessary Steps

As outlined earlier, the scientific and ethical justification for animal models of human diseases depends on their providing an opportunity to investigate disease biology and to determine potentially beneficial therapies for humans (Benatar, 2007). Thus, only after an animal model has proven to have satisfactory predictive value for humans, should it be refined as much as possible to reduce pain, suffering, distress, and lasting harm. If proven of no value, it should be abandoned. Such models should no longer receive regulatory approval nor funding, nor should they be accepted by scientific journals. SRs and meta-analyses of animal models as well as RAs of all animal experiments performed by independent experts would benefit animals and human patients, as they help to identify flawed studies and to eliminate misleading, invalid models, and experimental designs. Such a rigid quality control of animal-based research would most certainly lead to a significant reduction of animal use and, thus, to an increased effort to find more animal-free, robust, human biology-based models.

7.2 *Is the Biomedical Research Industry Shifting away from Animal Use?*

> The problem is that it hasn't worked, and it's time we stopped dancing around the problem. […] We need to refocus and adapt new methodologies for use in humans to understand disease biology in humans.
>
> ZERHOUNI, former head of the US National Institutes of Health, quoted in McManus, 2013

There is growing recognition that instead of focusing efforts on trying to refine animal experiments, a primary focus on human-relevant data is needed (Collins, 2011; Giri and Bader, 2015; Langley et al., 2015, 2017; Zerhouni, 2014), as a significant challenge that medical research is facing today is the understanding and possible treatment of chronic, complex diseases of which many are not well understood and, thus, cannot be modeled in other animals (Tsukamoto, 2016). Tsukamoto asks in a Drug Discovery Today editorial: "How can we replicate human diseases that develop later in life and/or result from a prolonged unhealthy lifestyle, far beyond the lifespan of rodent animals? What makes us expect that the outcomes from carefully controlled animal

experiments can be duplicated in patients with substantial heterogeneity across various aspects (age, gender, genetics, lifestyle, disease stage, etc.)?" Transgenic mice commonly used as disease models, oftentimes contain multiple copies of presumed disease-causing transgenes, and it is dubious "whether phenotypes seen in mice as a result of this 'genetic exaggeration' have any relevance to the corresponding human diseases" (Tsukamoto, 2016). Zerhouni (2014) calls for a new approach that redirects the drug-development paradigm that commences with the patient to explore the genetic foundation of molecular changes inherent to human pathophysiology.

As Russell and Burch remarked in 1959, "refinement is never enough, and we should always seek further for reduction and if possible replacement" (Chapter 4). Since 1959, we have gathered immense knowledge about animals and their consciousness, which has led to the public acknowledgment by a group of prominent neuroscientists that other animals are conscious too: The Cambridge Declaration on Consciousness (Low, 2012). Since 1959, the technology revolution has also immensely changed the field of life sciences and, hence, provides us with the tools to move away from using animals (Langley et al., 2015, 2017).

Current legislation, reflecting societal concerns, as well as the scientific failures of animal research should function to drive research, testing, and education away from using live animals. Some areas of education and training are already using animal-free teaching approaches, for ethical reasons and educational advances (see e.g., Bones et al., 2019, Chapter 23; Pawlowski et al., 2019, Chapter 22 in this Volume). In the area of chemical-toxicity testing, some progress has already been made in finding advanced non-animal methods, initiated, for example, through the pioneering *Toxicology in the 21st Century* (*Tox21*), a US federal initiative (National Research Council, 2007; National Toxicity Program, 2004; Rovida et al., 2015; Zurlo, 2012). However, the general tendency in toxicology is to introduce new methods without eradicating all the old (animal-based) ones (Rovida et al., 2015). Still, the acceptance of animal-free alternatives by regulators without additional animal-based tests, in the pharmaceutical and food-toxicity testing fields, should be possible when proven scientifically qualified for the specific context of use. However, awareness and acceptance of scientically-valid, non-animal methods is still low among regulators as well as research workers (Ramirez et al., 2015).

The high failure rate of drugs in the clinical phase (Begley and Ellis, 2012; Food and Drug Administration, 2004; Hutchinson and Kirk, 2011; Kola and Landis, 2004; Olson et al., 2000) indicates not only poor scientific quality and cognitive bias but also that animals are not good models for humans (e.g., Greek and Kramer, 2019, Chapter 17 in this Volume; Kramer and Greek, 2018; Knight, 2019, Chapter 14 in this Volume; Leist and Hartung, 2013); and the same

applies to food-safety testing in animals (Rovida et al., 2015). Already back in 2000, an eye-opening report (Olson et al., 2000) was published about the results of a multinational pharmaceutical company survey, which served to better understand the concordance of the toxicity of pharmaceuticals in humans compared with other animals. The weakness of animal studies to predict the human toxicity of drugs became apparent, as results revealed a human toxicity concordance rate of 71% when tested in multiple rodent as well as non-rodent species. When they compared humans with rodent species only, there was a 43% correlation; humans compared with non-rodent species showed a 63% match. Drug toxicity studies in animals are long-lasting and, hence, may cause severe suffering; and they are frequently not predictive for effects in humans (Hartung, 2009).

Cumulative knowledge is essential for scientific progress. Thus, there is increasing awareness of the importance of data sharing and collaboration to shift the paradigm away from using unsound animal models for drug toxicity testing. The human toxome project, a systematic mapping of the entirety of toxicity pathways, is ongoing in the area of chemical risk assessment. Rovida et al. (2015) suggested that this project should be extended to include the assessment of efficacy and adverse effects of drugs and food ingredients. Continued reliance on animal models appears implausible to enhance the current poor rate of clinical approval of new treatments. This is why Humane Society International initiated the Biomedical Research for the 21st Century (BioMed21) Collaboration. The BioMed21 Collaboration is working internationally with health experts, regulatory and research agencies, funding bodies, and others to develop innovative research roadmaps that focus on understanding human disease pathophysiology. The goal is to further this human-focused approach to studying, preventing, and treating disease (BioMed 21 Collaboration, n.d.). A central recommendation of the BioMed21 2015 workshop was to use the Adverse Outcome Pathway (AOP) concept in biomedical research. AOP, an important concept in toxicology, describes a logical sequence of causally-linked biological events that lead from the first action of a compound to an eventual adverse effect on human health (Langley et al., 2017). Furthermore, it was recommended that technological advances should be combined in human-specific tools and models. The importance of funding these new approaches was highlighted as well as the need for faster validation and acceptance by the scientific community, funding bodies, and scientific journals, who mostly still postulate the use of animals (Langley et al., 2015, 2017).

BioMed21 is a rare example for a non-animal-based approach in the area of applied research, which—together with the field of basic research—uses the majority of animals. Overall, there is little evidence that these fields are reducing the use of animals, as the 3Rs posit we must. Quite the contrary: Animal use

has been increasing in the new century (Taylor and Rego, 2016), mainly due to an increasing generation and use of genetically altered animals (Bailey, 2019, Chapter 19 in this Volume; Carbone, 2004; Ormandy, Schuppli and Weary, 2009; Ram, 2019, Chapter 15 in this Volume), which has, in recent years, been fueled by excitement over new technologies, such as CRISPR, an easier genetic modification technique that will most probably lead to a further steep increase in animal numbers and species modified (Bailey, 2019). These new technologies, however, have not kept their promise of improving translation between animal models and human health, as they have failed to increase the efficiacy and the safety of drugs (Hunter, 2011). For a detailed discussuion on the scientific and ethical issues of the genetic modification of animals, see Chapter 19 in this Volume (Bailey, 2019).

7.2.1 Funding

Progress in the development of replacement methods seems to be *limited most by the availability of funds*. Some governments and non-governmental organizations around the world are providing scarce funding, especially when compared to funds available for biomedical and life research as a whole. It is unclear how much of the annual worldwide funds—an estimated US$100 billion for biomedical research alone (Chalmers and Glasziou, 2009) and up to US$240 billion for all the life sciences (Røttingen et al., 2013)—are currently used for research centered around the use of animals, as it is not differentiated in the statistics (e.g., in Germany, BMBF, 2017). Daneshian (2016) estimated that in 2015, funds for projects with animals in Germany, including animal research facilities, were about €1920 million; funds for replacement methods ranged around €6.45 million. These financial means, mainly derived from German taxes, are distributed in opposition to Germany's declared political goal of working towards replacement of animal use at the national level (BMEL, 2015) as well as the EU level (European Parliament, 2010, Recital 10).

In preclinical human model development, the *Tissue Chips for Disease Modeling and Efficacy Testing* initiative, funded by US National Center for Advancing Translational Sciences (NCATS) of the National Institutes for Health (NIH), is a rare example. Its goal is to explore human microphysiological systems as potential facilitators of drug development in numerous disease areas. Its budget is approximately US$15 million, annually, for 13 two-year projects (NCATS, 2017); while NIH, being the biggest funder and research organization in the world, has annual funds of about US$39 billion for medical research alone (NIH, 2019). The EU framework program for research and innovation, Horizon 2020 (European Commission, n.d.), has, at the time of writing, supported 16 research projects devoted to alternative methods to animal testing, with a total of €90 million (European Parliament, 2017). The main research activities

are targeted towards developing complex *in silico* and *in vitro* human-based systems for better and more cost-effective safety and efficacy testing of chemicals, nanoparticles, vaccines, and drugs (European Parliament, 2017).

Between 1981 and 2015, the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung, BMBF) gave €160 million in funding for over 500 3Rs research projects. Aside from not exclusively funding replacement projects, the funds dedicated to the 3Rs were sparse; for example, in the 6-year period between 2010 and 2015, less than €20 million were available (BMBF, 2016). The UK National Centre of the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) is the largest funder of such research in the United Kingdom (Burden et al., 2015). Between 2004 and 2014, NC3Rs awarded 200 grants worth approximately US$54 million (Burden et al., 2015); the annual overall budget of NC3Rs is approximately €11.2 million (NC3Rs, n.d. d). In contrast, the German national 3Rs center, Zentrum zum Schutz der Versuchstiere (Bf3R), has an annual budget of €1.5 million to run all of its operations (Bundesinstitut für Risikobewertung, 2016) and provides approximately €350,000 to external replacement and refinement research groups per year (Bundesinstitut für Risikobewertung, 2018). Replacement research has to compete with refinement research for these limited funds (BMBF, 2016; NC3Rs, n.d. a).

A donor that exclusively provides money for the first R is the cosmetic company Lush, which in 2012 established the Lush Prize in collaboration with the UK not-for-profit group, Ethical Consumer Research Association (Redmond, 2019, Chapter 27 in this Volume). Lush provides £250,000 in funding each year for the main prize categories, with additional funds for regional awards in Asia and the Americas (Lush Prize, n.d.). An example for a charity providing some funding is People for the Ethical Treatment of Animals (PETA) International Science Consortium (PISC), which, in June 2017, awarded funding to develop four *in vitro* exposure systems to researchers from institutions in the United Kingdom, United States, and Belgium that are leaders in the development of non-animal methods to test the toxicity of airborne substances (PETA International Science Consortium, n.d. a). PISC assists with funding where promising *in vitro* or *in silico* techniques require further development or validation in order to gain regulatory acceptance. PISC focuses on toxicology and until, 2017, it has contributed about €2.9 million towards improving and implementing non-animal research methods (PETA International Science Consortium, n.d. b). The Alternatives Research & Development Foundation (ARDF) funds and promotes the development and validation of non-animal methods in biomedical research, product testing, and education and has provided US$3.25 million in funds since 1993 (Alternatives Research & Development Foundation, 2018).

The US National Anti-Vivisection Society (NAVS) provides some grants through the International Foundation for Ethical Research (IFER) for early career scientists to develop humane, human-relevant alternatives that replace animal use (NAVS, 2018). Overall, there are a few local and international initiatives and prizes but most focus on animal testing, while non-animal approaches in basic and applied research lag behind. Moreover, to ensure the field of animal-free, human-based research methods and approaches is continually and substantially growing, increased, stable governmental funding must be provided.

7.2.2 Education and Training

Another obstacle in shifting the current research paradigm is the limited availability of educational and training courses on animal-free methods and approaches in all areas of biomedical science, but especially in basic and applied research, since current available guidance documents and databases as well as courses almost exclusively focus on testing alternatives. There are some efforts being made to improve experimental design, conduct, and reporting; for example, online resources are available at some of the national 3Rs centers, such as at Norecopa, Norway's National Consensus Platform for the advancement of the 3Rs (Norecopa, 2016b) and the UK NC3Rs (NC3Rs, n.d. b, c), since quality issues of biomedical research has become apparent.

By EU law, the researcher must be well informed about state-of-the-art developments in the field of investigation, and animals must only be used if all possible alternatives are considered to be inadequate (EC Joint Research Centre, 2013). The EC Joint Research Centre's EU Reference Laboratory for Alternatives to Animal Testing – European Centre for the Validation of Alternative Methods (EURL ECVAM) Search Guide (EC Joint Research Centre, 2013) and Data Base Service on Alternative Methods to animal experimentation (DB-ALM) (EC Joint Research Centre, 2017) ought to assist with the search for alternatives to animal use. However, even for experts in the respective field, it is a lengthy and difficult task, as existing search systems do not support the necessary search strategies.

Altertox Academy, formerly CAAT Academy, offers hands-on training, but primarily for toxicologists, in human-relevant alternative methods and technologies (Altertox Academy, 2018). Education and training courses, mandatory for all animal researchers in the EU, include one animal-free methods module (e.g., FELASA B courses), but of a 40 hour FELASA B course, about one hour is dedicated to replacements, and generally only alternatives used in toxicology testing are covered (e.g., Berliner Kompaktkurse, 2017, p. 23). In 2016, the University of California (UC) San Diego offered a course that introduces participants to the available non-animal research methods, their efficacy, and how to

identify and implement them. It covered more areas than just regulatory toxicology (UC San Diego, 2018). However, detailed courses with extensive modules for all areas of the biomedical sciences currently do not exist.

7.2.3 Search Engine for Alternative Methods

What is urgently needed—aside from specific education and training courses—is an unbiased, freely available search engine that is able to find correlations regarding scientific purpose between animal experiments and alternative methods and, at the same time, 3Rs-relevant deviances in the methodologies (*in vitro* versus *in vivo*). Scientists from the Leibniz Institute for Social Sciences (GESIS) and the German Federal Institute for Risk Assessment (Bundesinstitut für Risikobewertung, BfR) have laid the foundation for such a search engine, using machine learning. The project is called SMAFIRA, which stands for "smart feature-based interactive ranking algorithm." The goal of SMAFIRA is to develop automated but mechanistically transparent search procedures that focus on such deviations and, thus, to provide an improved automatic support to search for non-animal methods (fisaonline, n.d.). This search engine will drastically reduce the number of documents scientists have to go through (GESIS, n.d.). A first version of the SMAFIRA search engine is anticipated to be available in the second half of 2019 (Daniel Butzke, BfR, personal communication, January 2019).

8 Ways to Work Towards Replacement

Directive 2010/63/EU, a progressive animal protection legislation in the field, sums up some important steps that have to be taken to work towards a paradigm shift, when it states (emphasis added): "The availability of alternative methods is highly dependent on the progress of the research into the development of alternatives. […] the Commission and the Member States should contribute through research and by other means to the **development and validation of alternative approaches**." (Recital 46). Article 47 declares: "The **Commission** and the **Member States shall contribute to the development and validation of alternative approaches** which could provide the same or higher levels of information as those obtained in procedures using animals […], and they **shall take such other steps as they consider appropriate to encourage research in this field**. […] Member States should, at national level, ensure the **promotion of alternative approaches and the dissemination of information** […]".

8.1 *Political Engagement*

The needed political engagement that Directive 2010/63/EU demands from its Member States to move towards an animal-free world of scientific experimentation was made a priority by the Dutch government in 2016. The Netherlands National Committee for the protection of animals used for scientific purposes (NCad) has developed a vision and plan of action for moving away from laboratory animal use. The Dutch goal is to phase out the utilization of animals in a number of fields by 2025, namely in regulatory testing of chemicals, food ingredients, pesticides and (veterinary) medicines, and biological products, such as vaccines (NCad, n.d.). The Committee also plans to steadily reduce animal involvement in regulatory preclinical research and basic research: "If we are to make the transition to non-animal research methods, we must make a paradigm shift away from existing mindsets and practices" (NCad, n.d., p. 3), a task which seems to be impossible without political involvement. The Dutch strategy holds the potential to act as a driver for other countries to follow this path.

8.2 *Legislative Change*

There is a need for regulators who are brave to move legislative change forward. The reason for the continued use of animals for regulatory testing is legislative, as existing policies require that new drug candidates are tested on animals before they can be assessed in human clinical trials, regardless of the fact that these animal tests are often unreliable in assessing safety and efficacy in humans (Greek and Kramer, 2019, Chapter 17 in this Voume). These regulations need to be amended according to scientific knowledge, and serious efforts need to be made to accelerate the development of advanced, humane, and human-relevant models (Archibald, Coleman and Drake, 2019, Chapter 18 in this Volume).

8.3 *Redeployment of Funds*

Absolutely essential for the paradigm change towards advanced, animal-free science and better healthcare is the redirection of funding. The limited funding for replacement research, oftentimes, has to compete with refinement research (e.g., BMBF, 2016; NC3Rs, n.d. a). These scarce funds should be used to further human biology-based approaches. Also, regarding taxpayers' money, the national governments, arguably, have the responsibility to use the funds in the name of a society that has repeatedly voiced that more needs to be done to replace animals in science. Moreover, our society is ethically evolving, with evidence of dwindling acceptance for animal suffering in the name of science. And it is being increasingly acknowledged that the continued reliance on animal models is unlikely to improve significantly the currently poor rate of

clinical approval of new treatments. Thus, animal-based research also contributes to resources being wasted (Harris, 2017; Ioannidis et al., 2014; Keen, 2019, Chapter 10 in this Volume).

Aside from redeploying funds, partially to preclinical human-relevant disease research (Langley et al., 2017) and to clinical rather than basic research (Pound and Bracken, 2014), a large part of funding should be dedicated to disease prevention efforts. To combat the increasing prevalence of dementia, for example, human-focused, non-animal models and methods, such as computational methods, advanced brain imaging techniques, and epidemiological studies should be given funding preference (Pistollato et al., 2016). Another extremely important area of disease prevention is basic public healthcare (Marks, 2012) as well as nutrition and lifestyle education. In addition, funds should also be used for pollution control, as pollution is currently found to be the largest environmental cause of disease and premature death around the world (Landrigan et al., 2017). The World Health Organization (WHO) estimated that around 3 million people die prematurely every year due to air pollution alone (Watts et al., 2017). In 2015, diseases caused by pollution were responsible for about 16% (9 Mio.) of all human deaths worldwide, which is three times more than deaths from tuberculosis, malaria, and AIDS combined and 15 times more than all wars and other means of violence together (Landrigan et al., 2017).

8.4 *Education and Training*

Education as well as re- and ongoing training about how to conduct *state-of-the-art science* and report it properly, as well as education on *research ethics* and *bioethics* are crucial. They enable students and scientists to gain a solid grounding in science based on non-animal models, while sincerely embracing the hierarchy of the 3Rs. Such learning objectives should be made available and should be mandatory for everyone planning to work or working in biomedical science. Education and retraining are the most important means to move away from the current thought culture and practice of animal use towards a new, humane research paradigm.

8.5 *Scientific Collaboration*

As Russell and Burch observed in 1959, "As we shall see, replacement is widely used in some fields, while in others it is very far from being exploited to the full, if at all. Moreover, such developments have been largely empirical, and largely independent of each other" (Chapter 5). At the moment, 3Rs experts are divided into replacement experts, on the one hand, and refinement experts, on the other. Animal welfare bodies and national committees in the EU (Directive 2010/63/EU, Recital 48), for example, are supposed to advise

scientists about the application of the 3Rs but seem to have little to no knowledge about available replacements and novel animal-free approaches to scientific questions (van Luijk et al., 2012; van Luijk et al., 2013). To achieve the ultimate goal in shifting the focus from refinement of animal use to replacement of animal use the animal research community needs to engage with replacement experts. National 3Rs centers should be equipped with a majority of experts in replacement methods, and a close collaboration between replacement experts and animal researchers appears crucial in moving towards animal replacement. To accelerate the development of new human biology-based approaches, a multidisciplinary approach is essential for bringing together the newest technologies and experts from various disciplines (Langely et al., 2017; Noor, 2019, Chapter 25 in this Volume).

9 Final Remarks

Looking into the future of animal-based science, Carbone (2004) wrote that morality and politics will continue to be the drivers for replacement research. Since the introduction of the principles, it has been widely held that animal researchers have an ethical responsibility to minimize any pain, distress, fear, suffering, and harm caused to animals when keeping them confined and utilizing them for invasive experiments without their consent. To apply the knowledge gained through animal welfare and refinement research is good veterinary and scientific practice, but it is not a substitute for reduction and replacement of animal experimentation. Indeed, Balls warned "that refinement can be used as a convenient way of showing commitment to the 3Rs, while ensuring that animal experimentation is seen as respectable and can be allowed to continue, while the fundamental ethical questions raised by it are avoided" (2010, p. 21). Thus, we have to be on guard that refinement is not used as a whitewashing tool, but its full application, which is an ethical imperative, must be guaranteed during the transition to human-relevant, animal-free methodologies.

Aside from extensive flaws in the way the majority of animals are housed and treated, and the poor conduct and reporting of many animal studies, the general lack of transparency around the use of animals in research as well as the low rate of critical appraisal of animal experiments are apparent. These failings have led to incorrect data and an overestimation of their significance (Cohen, 2018). Unnecessary harm inflicted upon these animals and, in the case of medical research, the harms done to patients who suffer from adverse reactions to drugs that were tested safe in animals or who are urgently waiting for treatments are serious issues that need to be addressed. A commitment to adhere to the 3Rs and to good scientific practice as well as to address societal

concerns about the use of animals in science would require a strong shift away from animals towards the use of human-relevant approaches.

References

Aarts, A.A., J.E. Anderson, C.J. Anderson, P.R. Attridge, A. Attwood and A. Fedor (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349(6251), pp. 1–8.

Abou-Ismail, U.A. and H.D. Mahboub (2011). The Effects of Enriching Laboratory Cages Using Various Physical Structures on Multiple Measures of Welfare in Singly-Housed Rats. *Laboratory Animals*, 45(3), pp. 145–153.

Akhtar, A.Z., J.J. Pippin and C.B. Sandusky (2009). Animal Studies in Spinal Cord Injury: A Systematic Review of Methylprednisolone. *Alternatives to Laboratory Animals*, 37(1), pp. 43–62.

AALAS (n.d.). AALAS National Meeting Abstract Archive. [online] Available at: https://www.aalas.org/national-meeting/abstract-archive [Accessed 13 August 2018].

Altertox Academy (2018). *Hands-on lab trainings and events in alternative methods in toxicology.* [online] Available at: http://academy.altertox.be [Accessed 12 August 2018].

Altman, D.G. (1994). The Scandal of Poor Medical Research. *British Medical Journal*, 308(6924), pp. 283–284.

Altman, D.G. (1980). Statistics and Ethics in Medical Research. Misuse of Statistics is Unethical. *British Medical Journal*, 281(6249), pp. 1182–1184.

Alternatives Research & Development Foundation (2018). *ARDF Annual Open Grant Program.* [online] Available at: http://www.ardf-online.org/ardf-grants.html [Accessed 12 August 2018].

André, V., C. Gau, A. Scheideler, J.A. Aguilar-Pimentel, O.V. Amarie, L. Becker, L. Garrett, W. Hans, S.M. Hölter, D. Janik and K. Moreth (2018). Laboratory Mouse Housing Conditions Can Be Improved Using Common Environmental Enrichment Without Compromising Data. *PLoS Biology*, 16(4), p. e2005019. [online] Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2005019 [Accessed 18 April 2018].

Anselme, P., M.J. Robinson and K.C. Berridge (2013). Reward Uncertainty Enhances Incentive Salience Attribution as Sign-tracking. *Behavioural Brain Research*, 238, pp. 53–61.

Archibald, K., R. Coleman and T. Drake (2019). Replacing animal tests to improve safety for humans. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill, pp. 417–442.

Augustsson, H., H.A. van de Weerd, C.L. Kruitwagen, V. Baumans (2003). Effect on Enrichment on Variation and Results in Light/Dark Test. *Laboratory Animals*, 37(4), pp. 328–340.

Bailey, J. (2018). Does the stress of laboratory life and experimentation on animals adversely affect research data? A critical review. *Alternatives to Laboratory Animals*, 46(5), pp. 291–305.

Bailey, J. (2019). Genetic modification of animals: scientific and ethical issues. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill pp. 443–479.

Bailoo, J.D., T.S. Reichlin and H. Würbel (2014). Refinement of Experimental Design and Conduct in Laboratory Animal Research. *ILAR Journal*, 55(3), pp. 383–391.

Baker, M. (2016). 1,500 Scientists Lift the Lid on Reproducibility. *Nature News*, 533(7604), p. 452.

Baker, D., K. Lidster, A. Sottomayor and S. Amor (2014). Two Years Later: Journals Are Not Yet Enforcing the ARRIVE Guidelines on Reporting Standards for Preclinical Animal Studies. *PLoS Biology*, 12(1), p. e1001756. [online] Available at: http://journals.plos.org/plosbiology/%20article?id=10.1371/journal.pbio.1001756 [Accessed 12 August 2018].

Balcombe, J. (2010). Laboratory rodent welfare: thinking outside the cage. *Journal of Applied Animal Welfare Science*, 13(1), pp. 77-88.

Balcombe, J.P., N.D. Barnard and C. Sandusky (2004). Laboratory Routines Cause Animal Stress. *Journal of the American Association for Laboratory Animal Science*, 43(6), pp. 42–51.

Baldwin, A.L., R.L. Primeau and W.E. Johnson (2006). Effect of Noise on the Morphology of the Intestinal Mucosa in Laboratory Rats. *Journal of the American Association for Laboratory Animal Science*, 45, pp. 74–82.

Balls, M. (1999). The biomedical sciences and the need for less inhumane animal procedures. In: C.F.M. Hendriksen and D.B. Morton, eds., *Humane Endpoints in Animal Experiments for Biomedical Research*. London: Royal Society of Medicine Press, pp. 1–4.

Balls, M. (2010). The Principles of Humane Experimental Technique: Timeless Insights and Unheeded Warnings. *Alternatives to Animal Experimentation*, 27(2), pp. 144–148. [online] Available at: http://www.altex.ch/resources/altex_2010_2_144_148_Balls.pdf [Accessed 12 August 2018].

Bath, P.M.W., L.J. Gray, A.J.G. Bath, A. Buchan, T. Miyata and A.R. Green (2009). Effects of NXY-059 in Experimental Stroke: An Individual Animal Meta-analysis. *British Journal of Pharmacology*,157(7), pp. 1157–1171.

Baumans, V. and P. van Loo (2013). How to Improve Housing Conditions of Laboratory Animals: The Possibilities of Environmental Refinement. *The Veterinary Journal*, 195(1), pp. 24–32.

Bayne, K. (2005). Potential for Unintended Consequences of Environmental Enrichment for Laboratory Animals and Research Results. *ILAR Journal*, 46(2), pp. 129–139.

Bayne, K. and H. Würbel (2014). The Impact of Environmental Enrichment on The Outcome Variability and Scientific Validity of Laboratory Animal Studies. *Revue*

Scientifique et Technique de l'OIE, 33(1), pp. 273–280. [online] Available at: https://www.oie.int/doc/ged/D13675.PDF [Accessed 12 August 2018].

Bayne, K., G.S. Ramachandra, E.A. Rivera and J. Wang (2015). The Evolution of Animal Welfare and the 3Rs in Brazil, China, and India. *Journal of the American Association for Laboratory Animal Science*, 54(2), pp. 181–191.

Begley, C.G. and L.M. Ellis (2012). Drug Development: Raise Standards for Preclinical Cancer Research. *Nature*, 483(7391), pp. 531–533.

Begley, C.G. and J.P.A. Ioannidis (2015). Reproducibility in Science. *Circulation Research*, 116(1), pp. 116–126.

Benatar, M. (2007). Lost in Translation: Treatment Trials in the SOD1 Mouse and in Human ALS. *Neurobiology of Disease*, 26(1), pp. 1–13.

Benefiel, A.C., W.K. Dong and W.T. Greenough (2005). Mandatory "Enriched" Housing of Laboratory Animals: The Need for Evidence-Based Evaluation. *ILAR Journal*, 46(2), pp. 95–105.

Bennett, E.L., M.C. Diamond, D. Krech and M.R. Rosenzweig (1964). Chemical and Anatomical Plasticity of Brain. *Science*, 146(3644), pp. 610–619.

Berliner Kompaktkurse (2017). *Laboratory Animal Science—Basic Course Main Focus: Mice/Rats*, p. 23. [online] Available at: https://www.berliner-kompaktkurse.de/assets/files/downloads/ph_kk_2017-01_web.pdf [Accessed 1 July 2017].

Bertolucci, L.F. (2011). Pandiculation: nature's way of maintaining the functional integrity of the myofascial system?. *Journal of Bodywork and Movement Therapies,* 15(3), pp. 268–280.

Bertrand, H.G.M.J., C. Sandersen and P.A. Flecknell (2018). Reported analgesic and anaesthetic administration to non-human primates undergoing experimental surgical procedure: 2010–2015. *Journal of Medical Primatology,* 47, pp. 217–225.

Beura, L., S. Hamilton, K. Bi, J. Schenkel, O. Odumade, K. Casey, E. Thompson, K. Fraser, P. Rosato, A. Filali-Mouhim, R. Sekaly, M. Jenkins, V. Vezys, W. Haining, S. Jameson and D. Masopust (2016). Normalizing the Environment Recapitulates Adult Human Immune Traits in Laboratory Mice. *Nature*, 532(7600), pp. 512–516.

BioMed21 Collaboration (n.d.). *About us*. [online] Available at: http://biomed21.org/about-us/ [Accessed on 23 May 2018].

BMBF (2016). *Alternativen zum Tierversuch. Informationen zur Förderung.* [online] Availabale at: https://www.bmbf.de/pub/Alternativmethoden_zum_Tierversuch.pdf [Accessed 1 August 2017].

Birke, L., Arluke, A. and Michael, M. (2007). *The Sacrifice: How Scientific Experiments Transform Animals and People*. Purdue University Press.

BMEL (2015). "*Mein langfristiges Ziel ist es, Tierversuche komplett zu ersetzen*." Bundesminister Schmidt eröffnet das Deutsche Zentrum zum Schutz von Versuchstieren. [online] Berlin, 25. September 2015. Available at: https://www.bmel.de/DE/Tier/Tierschutz/_texte/TierschutzTierforschung.html;jsessionid=DB4703F553CD731C7EA41D8F642C806E.2_cid367?docId=6671414 [Accessed 12 August 2018].

Bones, V.C., R.C.M. Garcia, G.G. Alves, R.L. Paixão, A.A. Rocha, K.V. Capilé and R. Bachinski (2019). Humane education: the tool for scientific revolution in Brazil. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Bracken, M.B. (2009). Why animal studies are often poor predictors of human reactions to exposure. *Journal of the Royal Society of Medicine*, 102(3), pp. 120–122.

Briel, M., K.F. Müller, J.J. Meerpohl, von E. Elm, B. Lang, E. Motschall, V. Gloy, F. Lamontagne, G. Schwarzer and D. Bassler (2013). Publication bias in animal research: A systematic review protocol. *Systematic Reviews*, 2, p. 23. [online] Available at: https://systematicreviewsjournal.biomedcentral.com/track/pdf/10.1186/2046-4053-2-23 [Accessed 12 August 2018].

Brown, M.J. (2014). Creating a culture of care. *NC3Rs News & Blog Online*. [online] Available at: https://www.nc3rs.org.uk/news/creating-culture-care [Accessed 12 August 2018].

Bundesinstitut für Risikobewertung (2016). *Fragen und Antworten zum Deutschen Zentrum zum Schutz von Versuchstieren* (*Bf3R*). [online] Available at: http://www.bfr.bund.de/cm/343/fragen-und-antworten-zum-deutschen-zentrum-zum-schutz-von-versuchstieren-bf3r.pdf [Accessed 12 August 2018].

Bundesinstitut für Risikobewertung (2018). *Bf3R Research Funding in the area of 3R – Replacement, Reduction and Refinement.* [online] Available at: https://www.bfr.bund.de/en/bf3r_research_funding_in_the_area_of_3r___replacement__reduction_and_refinement-62825.html [Accessed 12 August 2018].

Burden, N., K. Chapman, F. Sewell and V. Robinson (2015). Pioneering Better Science through the 3Rs: An Introduction to the National Centre for the Replacement, Refinement, and Reduction of Animals in Research (NC3Rs). *Journal of the American Association for Laboratory Animal Science*, *54*(2), pp. 198–208.

Burghardt, G.M. (1996). Environmental enrichment or controlled deprivation? In: G.M. Burghardt, J.T. Bielitzki, J.R. Boyce and D.O. Schaefer, eds., *The Well-being of Animals in Zoo and Aquarium Sponsored Research*. Greenbelt, MD: Scientists Center for Animal Welfare, pp. 91–101.

Burghardt, G.M. (1999). Deprivation and Enrichment in Laboratory Animal Environments, *Journal of Applied Animal Welfare Science*, 2(4), pp. 263–266.

Burn, C.C. (2017). Bestial Boredom: A Biological Perspective on Animal Boredom and Suggestions for Its Scientific Investigation. *Animal Behaviour*, 130, pp. 141–151.

Burn, C.C., A. Peters, M.J. Day and G.J. Mason (2006). Long-Term Effects of Cage-Cleaning Frequency And Bedding Type on Laboratory Rat Health, Welfare, and Handleability: A Cross-Laboratory Study. *Laboratory Animals*, 40(4), pp. 353–370.

Cabib, S. (2006). The Neurobiology of stereotypy II: The role of stress. In: G. Mason and J. Rushen, eds., *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare*, 2nd ed. Oxfordshire, UK: CAB International.

CAMARADES (2014). *Bringing Evidence to Translational Medicine*. [online] Available at: http://www.dcn.ed.ac.uk/camarades/default.htm [Accessed 12 August 2018].

Campi, K.L., C.E. Collins, W.D. Todd, J. Kaas and L. Krubitzer (2011). Comparison of Area 17 Cellular Composition in Laboratory and Wild-Caught Rats Including Diurnal and Nocturnal Species. *Brain, Behavior and Evolution*, 77, pp. 116–130.

Cao, L., X. Liu, E. Lin, C. Wang, E. Choi, V. Riban, B. Lin and M. During (2010). Environmental and Genetic Activation of a Brain-Adipocyte BDNF/Leptin Axis Causes Cancer Remission and Inhibition. *Cell*, 142(1), pp. 52–64.

Carbone, L. (2004). What Animals Want: Expertise and Advocacy in Laboratory Animal Welfare Policy. New York: Oxford University Press.

Carbone, L. and J. Austin (2016). Pain and Laboratory Animals: Publication Practices for Better Data Reproducibility and Better Animal Welfare. *PLoS One*, 11(5). [online] Avaiable at: http://journals.plos.org/plosone/article/related?id=10.1371/journal.pone.0155001 [Accessed 12 August].

Cavanaugh, S.E., Pippin, J.J. and Barnard, N.D. (2014). Animal Models of Alzheimer Disease: Historical Pitfalls and a Path Forward. Alternatives to Animal Experimentation, 31(3), pp. 279–302.

Center for Open Science (n.d.). *Our Mission is to Increase Openness, Integrity, and Reproducibility of Research*. [online] Available at: https://cos.io/about/mission/ [Accessed 12 August 2018].

Chalmers, I. and P. Glasziou (2009). Avoidable Waste in the Production and Reporting of Research Evidence. *Obstetrics & Gynecology*, 114(6), pp. 1341–1345.

Charbonneau, R., L. Niel, E. Olfert, M. von Keyserlingk and C. Griffin (2010). CCAC Guidelines on: Euthanasia of Animals Used in Science. *Canadian Council on Animal Care.* Ottawa ON, Canada, pp. 1–32. [online] Available at: https://www.ccac.ca/Documents/Standards/Guidelines/Euthanasia.pdf [Accessed 12 August 2018].

Chawla, S.D. (2017). 'One-size-fits-all' threshold for *P* values under fire. Scientists hit back at a proposal to make it tougher to call findings statistically significant. *Nature News*. [online] Available at: https://www.nature.com/news/one-size-fits-all-threshold-for-p-values-under-fire-1.22625?WT.mc_id=TWT_NatureNews [Accessed 12 August 2018].

Chesler, E.J., S.G. Wilson, W.R. Lariviere, S.L. Rodriguez-Zas and J.S. Mogil (2002). Influences of Laboratory Environment on Behavior. *Nature Neuroscience,* 5, pp. 1101–1102.

Clarkson, J.M., D.M. Dwyer, P.A. Flecknell, M.C. Leach and C. Rowe (2018). Handling method alters the hedonic value of reward in laboratory mice. *Scientific Reports*, 8(1), p. 2448. [online] Available at: https://www.nature.com/articles/s41598-018-20716-3 [Accessed 12 August 2018].

Clemence, M. and J. Leaman (2016). *Public Attitudes to Animal Research in 2016*. A report by Ipsos MORI for the Department for Business, Energy & Industrial Strategy, Ipsos MORI Social Research Instititute. [online] Available at: https://www

.ipsos.com/sites/default/files/publication/1970-01/sri-public-attitudes-to-animal-research-2016.pdf [Accessed 12 August 2018].

Cohen, D. (2018). Oxford TB Vaccine Study Calls into Question Selective Use of Animal Data. *British Medical Journal*, 360, p. j5845.

Collins, F.S. (2011). Reengineering Translational Science: The Time Is Right. *Science Translational Medicine*, 3(90), p. 90cm17. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5101940/ [Accessed 12 August 2018].

Collins, F.S. and L.A. Tabak (2014). NIH Plans To Enhance Reproducibility. *Nature,* 505, pp. 612–613.

Commission of the European Communities (2007). Recommendation of 18 June 2007 on Guidelines for the accommodation and Care of Animals Used for Experimental and Other Scientific Purposes (2007/526/EC). *Official Journal of the European Communities*, L197, pp. 1–89. [online] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007H0526&from=EN [Accessed 15 August 2017].

Coulter, C., P. Flecknell and C. Richardson (2009). Reported Analgesic Administration to Rabbits, Pigs, Sheep, Dogs and Non-Human Primates Undergoing Experimental Surgical Procedures. *Laboratory Animals*, 43(3), pp. 232–238.

Coulter, C., P. Flecknell, M. Leach and C. Richardson (2011). Reported Analgesic Administration To Rabbits Undergoing Experimental Surgical Procedures. *BMC Veterinary Research*, 7(1), p. 12. [online] Available at: https://doi.org/10.1186/1746-6148-7-12 [Accessed 12 August 2018].

Council of Europe (1986). *Convention for the Protection of Vertebrate Animals Used for Experimental and other Scientific Purposes.* European Treaty Series No. 123. [online] Available at: https://rm.coe.int/168007a67b [Accessed 12 August 2018].

Crabbe, J.C., D. Wahlsten and B.C. Dudek (1999). Genetics of Mouse Behavior: Interactions with Laboratory Environment. *Science*, 284(5420), pp. 1670–1672.

Daneshian, M. (2016). Vergleiche der Förderungen von Alternativen in Deutschland – Europa – USA. In: *Forschung ohne Tierversuche – verliert Deutschland den Anschluss? Zum Weltaktionstag gegen Tierversuche.* Talk, Berlin. [online] Program available at: https://www.urania.de/forschung-ohne-tierversuche-verliert-deutschland-den-anschluss [Accessed 6 August 2017].

Daneshian, M., F. Busquet, T. Hartung and M. Leist (2015). Animal Use for Science in Europe. *Alternatives to Animal Experimentation,* 32, pp. 261–274.

Dawson, E., T. Gilovich and D.T. Regan (2002). Motivated Reasoning and Performance on the was on Selection Task. *Personality and Social Psychology Bulletin*, 28(10), pp. 1379–1387.

de Vries, R.B., P. Buma, M. Leenaars, M. Ritskes-Hoitinga and B. Gordijn (2012). Reducing the Number of Laboratory Animals Used in Tissue Engineering Research by Restricting the Variety of Animal Models. Articular Cartilage Tissue Engineering as a Case Study. *Tissue Engineering Part B: Reviews*, 18(6), pp. 427–435.

Dirnagl, U. and M. Fisher (2012). International, Multicenter Randomized Preclinical Trials in Translational Stroke Research: It's Time To Act. *Journal of Cerebral Blood Flow & Metabolism*, 32, pp. 933–935.

During, M.J., X. Liu, W. Huang, D. Magee, A. Slater, T. McMurphy, C. Wang and L. Cao (2015). Adipose VEGF Links the White-To-Brown Fat Switch with Environmental, Genetic, and Pharmacological Stimuli in Male Mice. *Endocrinology,* 156, pp. 2059–2073.

EC Expert Working Group for Project Evaluation and Retrospective Assessment (2013). 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. [online] Available at: http://ec.europa.eu/environment/chemicals/lab_animals/pdf/Endorsed_PE-RA.pdf [Accessed 12 August 2017].

EC Joint Research Centre (2013). *The EURL ECVAM Search Guide. Good Search Practice on Animal Alternatives*. EUR 24391 EN, Re-edition. [online] Available at: https://ec.europa.eu/jrc/en/publication/books/eurl-ecvam-search-guide-good-search-practice-animal-alternatives [Accessed 12 August 2017].

EC Joint Research Centre (2017). *EURL ECVAM Data Base Service on Alternative Methods to Animal Experimentation* (DB-ALM). [online] Availble at: https://ecvam-dbalm.jrc.ec.europa.eu/methods-and-protocols [Accessed 12 August 2018].

Eisen, J.A., E. Ganley and C.J. MacCallum (2014). Open Science and Reporting Animal Studies: Who's Accountable? *PLoS Biology,* 12(1), p. e1001757. [online] Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001757 [Accessed 12 August 2018].

Enserink, M. (2017). Sloppy reporting on animal studies proves hard to change. *Science*, 357(6358), pp. 1337–1338.

Errington, T.M., E. Iorns, W. Gunn, F.E. Tan, J. Lomax and B.A. Nosek (2014). An Open Investigation of the Reproducibility of Cancer Biology Research. *Elife*, 3, p. e04333.

Eskola, S., M. Lauhikari, H.M. Voipio, M. Laitinen and T. Nevalainen (1999). Environmental Enrichment May Alter the Number of Rats Needed To Achieve Statistical Significance. *Scandinavian Journal of Laboratory Animal Science,* 26, pp. 134–144.

European Citizen's Initiative (2016). *Stop Vivisection.* [online] Available at: http://www.stopvivisection.eu/de [Accessed 12 August 2018].

European Commission (2013). Report from the Commission to the Council and the European Parliament. Seventh Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union. *COM (2013) 859 final.* [online] Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:e99d2a56-32fc-4f60-ad69-61ead7e377e8.0001.03/DOC_1&format=PDF [Accessed 12 August 2018].

European Commission (2014). National Competent Authorities for the implementation of Directive 2010/63/EU on the Protection of Animals Used for Scientific Purposes.

A working document on the development of a common education and training framework to fulfill the requirements under the Directive. [online] Available at: http://ec.europa.eu/environment/chemicals/lab_animals/pdf/Endorsed_E-T.pdf [Accessed 15 July 2018].

European Commission (n.d.). *Horizon2020. The EU Framework Programme for Research and Innovation*. [online] Availble at: https://ec.europa.eu/programmes/horizon2020/en/find-your-area [Accessed 12 August 2018].

European Parliament (2010). 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. *Official Journal of the European Communities*, L276, pp. 33–79. [online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32010L0063 [Accessed 23 July 2017].

European Parliament (2017). *Parliamentary questions: Question Reference E-004265/2017*. Answer given by Mr Moedas on behalf of the Commission. [online] Available at: http://www.europarl.europa.eu/doceo/document/E-8-2017-004265-ASW_EN.html [Accessed 12 August 2018].

EUSAAT (2015). *19th European Congress on Alternatives to Animal Testing, 16th Annual Congress of EUSAAT*. Alternatives to Animal Experimentation Proceedings, 4(2). [online] Available at: http://www.eusaat-congress.eu/images/2015/Abstractbook_EUSAAT_2015_Linz_2015.pdf [Accessed 12 August 2018].

FELASA (2016). *Detailed program FELASA 2016*. [online] Available at: http://felasa2016.eu/wp-content/uploads/2016/06/DETAILED_PROGRAMME_FELASA_2016_V3.6.pdf [Accessed 12 August 2018].

fisaonline (n.d.). SMAFIRA – Smart Feature-based interactive ranking (SMAFIRA) Project. [online]. Available at: https://fisaonline.de/en/find-institutions/federal-research-institutions/?tx_fisaresearch_federalresearchinstitutions%5Bp_id%5D=4377&tx_fisaresearch_federalresearchinstitutions%5Baction%5D=projectDetails&tx_fisaresearch_federalresearchinstitutions%5Bcontroller%5D=Projects&cHash=1223e8b07c976b94f16b67e2f0e4af1b#more [Accessed 23 Janaury 2018].

Flecknell, P.A. (2016). *Laboratory Animal Anaesthesia*. 4th ed., London: Academic Press.

Flecknell, P.A. (2018). Rodent Analgesia: Assessment and Therapeutics. *The Veterinary Journal,* 232, pp. 70–77.

Fonseca, K., P. Rodrigues, I. Olsson and M. Saraiva (2017). Experimental Study Of Tuberculosis: From Animal Models To Complex Cell Systems and Organoids. *PLoS Pathogens*, 13(8) [online] Available at: https://doi.org/10.1371/journal.ppat.1006421 [Accessed 12 August 2018].

Food and Drug Administration (2018). *Food and Drug Administration Modernization Act* (*FDAMA*) [online] Available at: https://www.fda.gov/RegulatoryInformation/LawsEnforcedbyFDA/SignificantAmendmentstotheFDCAct/FDAMA/default.htm [Accessed 12 August 2018].

Food and Drug Administration (2004). *Innovation or Stagnation? Challenge and Opportunity on the Critical Path to New Medical Products*. US Department of Health and Human Services, Food and Drug Administration, pp. 1–32. [online] Available at: http://www.who.int/intellectualproperty/documents/en/FDAproposals.pdf [Accessed 12 August 2018].

Franco, N.H., Sandøe, P. and Olsson, I.A.S. (2018). Researchers' attitudes to the 3Rs—An upturned hierarchy?. *PloS One*, 13(8), p. e0200895. [online] Available at: https://doi.org/10.1371/journal.pone.0200895 [Accessed 17 August 2018].

Freedman, L.P., I.M. Cockburn, T.S. Simcoe (2015). The Economics of Reproducibility in Preclinical Research. *PLoS Biology,* 13(6), p. e1002165. [online] Available at: https://doi.org/10.1371/journal.pbio.1002165 [Accessed 15 July 2017].

Fureix, C., M. Walker, L. Harper, K. Reynolds, A. Saldivia-Woo and G. Mason (2016). Stereotypic Behaviour in Standard Non-enriched Cages Is an Alternative To Depression-like Responses in C57BL/6 Mice. *Behavioural Brain Research*, 305, pp. 186–190.

Garner, J.P. (2005). Stereotypies and Other Abnormal Repetitive Behaviors: Potential Impact on Validity, Reliability, and Replicability of Scientific Outcomes. *ILAR Journal*, 46(2), pp. 106–117.

Garner, J.P. and G.J. Mason (2002). Evidence for a Relationship Between Cage Stereotypies and Behavioural Disinhibition in Laboratory Rodents. *Behavioural Brain Research*, 136(1), pp. 83–92.

Gärtner, K. (1999). Cage Enrichment Occasionally Increases Deviation of Quantitative Traits. In: *International Joint Meeting 12th ICLAS General Assembly and Conference and 7th FELASA Symposium.* Madrid: SECAL, pp. 207–210.

Gaskill, B. and J. Garner (2014). Letter-To-The-Editor on "Not so hot: Optimal Housing Temperatures for Mice To Mimic the Thermal Environment of Humans". *Molecular Metabolism*, 3(4), pp. 335–336. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4060287/ [Accessed 12 August 2017].

Gaskill, B., C. Gordon, E. Pajor, J. Lucas, J. Davis and J. Garner (2012). Heat or Insulation: Behavioral Titration of Mouse Preference for Warmth or Access To a Nest. *PLoS One*, 7(3), p.e32799. [online] Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0032799 [Accessed 12 August 2017].

Gaskill, B. and K. Pritchett-Corning (2016). Nest Building as an Indicator of Illness in Laboratory Mice. *Applied Animal Behaviour Science*, 180, pp. 140–146.

Gaskill, B., S. Rohr, E. Pajor, J. Lucas and J. Garner (2009). Some Like It Hot: Mouse Temperature Preferences in Laboratory Housing. *Applied Animal Behaviour Science*, 116(2–4), pp. 279–285.

Gaskill, B., C. Winnicker, J. Garner and K. Pritchett-Corning (2013). The Naked Truth: Breeding Performance in Nude Mice with and without Nesting Material. *Applied Animal Behaviour Science*, 143(2–4), pp. 110–116.

GESIS (n.d.). SMAFIRA Computer-aided modeling of signatures and features of relevant scientific content. Available at: https://www.gesis.org/en/research/external-funding-projects/archive/smafira/ [Accessed 17 August 2017].

Giri, S. and A. Bader (2015). A Low-Cost, High-quality New Drug Discovery Process Using Patient-derived Induced Pluripotent Stem Cells. *Drug Discovery Today*, 20(1), pp. 37–49.

Glasziou, P., D.G. Altman, P. Bossuyt, I. Boutron, M. Clarke, S. Julious, S. Michie, D. Moher and E. Wager (2014). Reducing Waste from Incomplete or Unusable Reports of Biomedical Research. *The Lancet*, 383(9913), pp. 267–276.

Gluck, J.P. (2016). Voracious Science & Vulnerable Animals. A Primate Scientist's Ethical Journey. The University of Chicago Press: Chicago and London.

Gordon, C.J. (1993). *Temperature regulation in laboratory rodents*. New York: Cambridge University Press.

Gordon, C.J. (2012). Thermal Physiology of Laboratory Mice: Defining Thermoneutrality. *Journal of Thermal Biology*, 37(8), pp. 654–685.

Gouveia, K. and J.L. Hurst (2013). Reducing Mouse Anxiety During Handling: Effect of Experience with Handling Tunnels. *PLoS One*, 8(6), p. e66401.

Gouveia, K. and J.L. Hurst (2017). Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. *Scientific Reports*, 7, p. 44999, [online] Available at: https://www.nature.com/articles/srep44999 [Accessed 12 August 2018].

Greek, R. and L.A. Kramer (2019). The scientific problems with using non-human animals to predict human response to drugs and disease. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Gross, A., S. Richter, A. Engel and H. Würbel (2012). Cage-induced Stereotypies, Perseveration, and the Effects of Environmental Enrichment in Laboratory Mice. *Behavioural Brain Research*, 234(1), pp. 61–68.

Gruen, L. (2014). *The Ethics of Captivity*. New York: Oxford University Press.

Gurevitch, J., J. Koricheva, S. Nakagawa and G. Stewart (2018). Meta-analysis and the Science of Research Synthesis. *Nature*, 555(7695), pp. 175–182.

Gurfein, B.T., O. Davidenko, M. Premenko-Lanier, J.M. Milush, M. Acree, M.F. Dallman, C. Touma, R. Palme, V.A. York, G. Fromentin, N. Darcel, D.F. Nixon and F.M. Hecht (2014).Environmental Enrichment Alters Splenic Immune Cell Composition and enhances Secondary Influenza Vaccine Responses in Mice. *Molecular Medicine,* 20(1), pp. 179–190.

Hair, K., Macleod, M. R., Sena, E. S. and IICARus Collaboration (2018). A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). bioRxiv, p. 370874.

Hartshorne, J. and A. Schachner (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience*, 6, p. 8.

[online] Available at: https://www.frontiersin.org/articles/10.3389/fncom.2012.00008/full [Accessed 12 August 2018].

Hartung, T. (2009). Toxicology for the twenty-first century. *Nature*, 460(7252), pp. 208–212.

Hartung, T. (2013). Food for Thought Look Back in Anger—What Clinical Studies Tell Us About Preclinical Work. *Alternatives to Animal Experimentation*, 30(3), pp. 275–291. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3790571/ [Accessed 12 August 2018].

Harris, R. (2017). *Rigor mortis. How sloppy science creates worthless cures, crushes hope, and wastes billions*. New York: Basic.

Hau, J. (2008). Animal models for human disease: An overview. In: P.M. Conn, ed., *Sourcebook of Models for Biomedical Research*. Totowa: Humana Press Inc., pp. 3–9.

Hawkins, P., R. Armstrong, T. Boden, P. Garside, K. Knight, E. Lilley, M. Seed, M. Wilkinson and R. Williams (2015). Applying Refinement to the Use of Mice and Rats in Rheumatoid Arthritis Research. *Inflammopharmacology*, 23(4), pp. 131–150.

Hebb, D.O. (1947). The Effects of Early Experience on Problem Solving at Maturity. *American Psychologist*, 2, pp. 306–307.

Hendriksen, C., D. Morton and K. Cussler (2011). Use of humane end points to minimise suffering. In: B. Howard, T. Nevalainen and G. Perretta, eds., *The Cost Manual of Laboratory Animal Care and Use*. Florida: CRC Press, pp. 333–353.

Hennessy, M.B., McCowan, B., Jiang, J. and Capitanio, J.P. (2014). Depressive-like behavioral response of adult male rhesus monkeys during routine animal husbandry procedure. *Frontiers in Behavioral Neuroscience*, 8, p. 309. [online] Available at: https://www.frontiersin.org/articles/10.3389/fnbeh.2014.00309/full [Accessed 17 August 2018].

Herrmann, K. (2013). Directive 2010/63/EU—A Chance for More Humane Education?. In: *12th FELASA 2013 SECAL Congress "Animal Research: Better Science with Fewer Animals".* Barcelona: SECAL, Abstracts of scientific papers 12th FELASA SECAL Congress, Journal of the American Association for Laboratory Animal Science, 52(3), p. 389. [online] Available at: https://www.researchgate.net/publication/282816209_jaalas_FELASA_2013 [Accessed 12 August 2018].

Herrmann, K. and P.A. Flecknell (2018a). Retrospective review of anesthetic and analgesic regimens used in animal research proposals. *Alternatives to Animal Experimentation,* 36(1), pp. 65–80. [online] Available at: https://www.altex.org/index.php/altex/article/view/780/1234 [Accessed 14 September 2018].

Herrmann, K. and Flecknell, P.A. (2018b): Application of humane endpoints and humane killing methods in animal research proposals – a retrospective review. *Alternatives to Laboratory Animals*, 46(6), pp. 1–17.

Herrmann, K. and Flecknell, P.A. (2018c): Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals – a retrospective review. *Alternatives to Laboratory Animals*, 46(5), pp. 273–289.

Herrmann, K., K. Köpernik and M. Biedermann (2009). Ein Leitfaden für die Teilprüfung der "Unerlässlichkeit" im Hinblick auf "Refinement". In: D. Borchers and J. Luy, eds., *Der Ethisch Vertretbare Tierversuch, Kriterien und Grenzen*, Padaborn: mentis, pp. 219–234.

Herrmann, K. and H. Ratsch (2010): Bessere Haltungsbedingungen für Labortiere: Überwachung von Versuchstierhaltungen – die Frage des Enrichments. *Deutsches Tierärzteblatt*, 4, pp. 492–499.

Herzog, H. (2002). Ethical Aspects of Relationships Between Humans and Research Animals. *ILAR Journal*, 43(1), pp. 27–32.

Hess, S.E., S. Rohr, B.D. Dufour, B.N. Gaskill, E.A. Pajor and J.P. Garner (2008). Home Improvement: C57BL/6J Mice Given More Naturalistic Nesting Materials Build Better Nests. *Journal of the American Association for Laboratory Animal Science*, 47(6), pp. 25–31.

Higgins, J.P., D.G. Altman, P.C. Gøtzsche, P. Jüni, D. Moher, A.D. Oxman, J. Savović, K.F. Schulz, L. Weeks and J.A. Sterne (2011). The Cochrane Collaboration's Tool for Assessing Risk of Bias in Randomised Trials. *British Medical Journal*, 343, p. d5928.

Hooijmans, C.R., J. IntHout, M. Ritskes-Hoitinga and M.M. Rovers (2014a). Meta-analyses of Animal Studies: An Introduction of a Valuable Instrument To Further Improve Healthcare. *ILAR journal*, 55(3), pp. 418–426. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4276598/ [Accessed 15 July 2017].

Hooijmans, C.R., M. Leenaars and M. 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), pp. 167–182. [online] Available at: http://repository.ubn.ru.nl/bitstream/handle/2066/89153/89153.pdf [Accessed 15 August 2017].

Hooijmans, C.R. and M. Ritskes-Hoitinga (2013). Progress in Using Systematic Reviews of Animal Studies To Improve Translational Research. *PLoS Medicine*, *10*(7), p. e1001482. [online] Available at: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001482 [Accessed on 14 November 2017].

Hooijmans, C.R., M.M. Rovers, R.B. de Vries, M. Leenaars, M. Ritskes-Hoitinga and M.W. Langendam (2014b). SYRCLE's Risk of Bias Tool for Animal Studies. *BMC Medical Research Methodology*, 14(1), p. 43. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4230647/ [Accessed on 14 November 2017].

Howerton, C.L., J.P. Garner and J.A. Mench (2008). Effects of a Running Wheel-igloo Enrichment on Aggression, Hierarchy Linearity, and Stereotypy in Group-housed Male CD-1 (ICR) Mice. *Applied Animal Behaviour Science*, 115(1), pp. 90–103.

Hunter, A.J. (2011). Have Animal Models of Disease Helped or Hindered the Drug Discovery Process. *Annals of the New York Academy of Sciences*, 1245, pp. 1–2.

Hurst, J.L. and R.S. West (2010). Taming Anxiety in Laboratory Mice. *Nature Methods*, 7, pp. 825–826. [online] Available at: https://www.nature.com/nmeth/journal/v7/n10/full/nmeth.1500.html [Accessed on 14 January 2017].

Hutchinson, L. and R. Kirk (2011). High Drug Attrition Rates—Where Are We Going Wrong?. *Nature Reviews Clinical Oncology*, 8(4), pp. 189–190.

Ioannidis, J.P.A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8), p. e124. [online] Available at: https://doi.org/10.1371/journal.pmed.0020124 [Accessed 15 July 2017].

Ioannidis, J.P.A. (2012). Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), pp. 645–654. [online] Available at: http://journals.sagepub.com/doi/full/10.1177/1745691612464056 [Accessed 15 July 2017].

Ioannidis, J.P.A., S. Greenland, M.A. Hlatky, M.J. Khoury, M.R. Macleod, D. Moher, K.F. Schulz and R. Tibshirani (2014). Increasing Value and Reducing Waste in Research Design, Conduct, and Analysis. *The Lancet*, 383(9912), pp. 166–175.

Jirkof, P. (2015). Effects of Experimental Housing Conditions on Recovery of Laboratory Mice. *Lab Animal*, 44(2), pp. 65–70.

Joffe, A.R., M. Bara, N. Anton and N. Nobis (2016). Expectations for the Methodology and Translation Of Animal Research: A Survey of the General Public, Medical Students, and Animal Researchers in North America. *Alternatives to Laboratory Animals*, 44(4), pp. 361–381.

Johnson, E.A., D.S. Sharp and D.B. Miller (2000). Restraint as a Stressor in Mice: Against the Dopaminergic Neurotoxicity of d-MDMA, Low Body Weight Mitigates Restraint-induced Hypothermia and Consequent Neuroprotection. *Brain Research*, 875(1), pp. 107–118.

Johnson, J. and A. Smajdor (2019). Human wrongs in animal research—A focus on moral injury and reification. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Jones, J.M. (2017). Americans Hold Record Liberal Views on Most Moral Issues. *Gallup Poll Social Series*. [online] Available at: http://www.gallup.com/poll/210542/americans-hold-record-liberal-views-moral-issues.aspx?g_source=2017+poll+animals&g_medium=search&g_campaign=tiles [Accessed 10 June 2017].

Justice, M.J. and P. Dhillon (2016). Using the Mouse to Model Human Disease: Increasing Validity and Reproducibility. *Disease Model Mechanisms*, 9(2), pp. 101–103.

Kaplan, R.M. and V.L. Irvin (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased Over Time. *PLoS One*, 10(8), p. e0132382. [online] Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132382 [Accessed 10 June 2017].

Karp, C.L. (2012). Unstressing Intemperate Models: How Cold Stress Undermines Mouse Modeling. *Journal of Experimental Medicine*, 209(6), pp. 1069–1074.

Keen, J. (2019). Wasted money in US biomedical and agricultural animal research. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Kempermann, G., H.G. Kuhn and F.H. Gage (1997). More Hippocampal Neurons in Adult Mice Living in an Enriched Environment. *Nature*, 386(6624), p. 493.

Kerr, N.L. (1998). HARKing: Hypothesizing After the Results Are Known. *Personality and Social Psychology Review*, 2(3), pp. 196–217.

Kilkenny, C., W.J. Browne, I.C. Cuthill, M. Emerson and D.G. Altman (2010). Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLoS Biology*, 8(6), p. e1000412. [online] Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000412 [Accessed 14 May 2017].

Klein, H. and K. Bayne (2007). Establishing a Culture of Care, Conscience, and Responsibility: Addressing the Improvement of Scientific Discovery and Animal Welfare Through Science-based Performance Standards. *ILAR Journal*, 48(1), pp. 3–11.

Knight, A. (2008). 127 Million Non-Human Vertebrates Used Worldwide for Scientific Purposes in 2005. *Alternatives to Laboratory Animals*, 36(5), pp. 494–496.

Knight, A. (2019). Critically evaluating animal research. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill pp. 321–340.

Kola, I. and J. Landis (2004). Can the Pharmaceutical Industry Reduce Attrition Rates? *Nature Reviews Drug Discovery,* 3(8), p. 711–715.

Kramer, L.A. and R. Greek (2018). Human Stakeholders and the Use of Animals in Drug Development. *Business and Society Review,* 123(1), pp. 3–58.

Krech, D., M.R. Rosenzweig and E.L. Bennett (1960). Effects of Environmental Complexity and Training on Brain Chemistry. *Journal of Comparative and Physiological Psychology,* 53, pp. 509–514.

Lahvis, G.P. (2017). Unbridle Biomedical Research From the Laboratory Cage. *Elife*, 6. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5503508/ [Accessed 4 July 2017].

Lancaster, M.A. and J.A. Knoblich (2014). Generation of Cerebral Organoids From Human Pluripotent Stem Cells. *Nature Protocols*, 9(10), pp. 2329–2340.

Landis, S.C., S.G. Amara, K. Asadullah, C.P. Austin, R. Blumenstein, E.W. Bradley, R.G. Crystal, R.B. Darnell, R.J. Ferrante, H. Fillit and R. Finkelstein (2012). A Call for Transparent Reporting To Optimize The Predictive Value of Preclinical Research. *Nature*, 490(7419), pp. 187–191.

Landrigan, P.J., R. Fuller, N.J. Acosta, O. Adeyi, R. Arnold, A.B. Baldé, R. Bertollini, S. Bose-O, J.I. Boufford, P.N. Breysse and T. Chiles (2017). The Lancet Commission on Pollution and Health. *The Lancet*, pp. 462–512.

Langley, G., I. Adcock, F. Busquet, K. Crofton, E. Csernok, C. Giese, T. Heinonen, K. Herrmann, M. Hofmann-Apitius, B. Landesmann, L. Marshall, E. McIvor, A. Muotri, F. Noor, K. Schutte, T. Seidle, A. van de Stolpe, H. Van Esch, C. Willett and G. Woszczek (2017). Towards a 21st Century Roadmap for Biomedical Research and Drug Discovery: Consensus Report and Recommendations. *Drug Discovery Today*, 22(2), pp. 327–339.

Langley, G., C.P. Austin, A.K. Balapure, L.S. Birnbaum, J.R. Bucher, J. Fentem, S.C. Fitzpatrick, J.R. Fowle III, R.J. Kavlock, H. Kitano and B.A. Lidbury (2015). Lessons from

Toxicology: Developing a 21st Century Paradigm for Medical Research. *Environmental Health Perspectives*, 123(11), p. A268–A272.

Leary, S., W. Underwood, R. Anthony, S. Cartner, D. Corey, T. Grandin, C. Greenacre, S. Gwaltney-Brant, M.A. McCrakin, R. Meyer and D. Miller (2013). *AVMA Guidelines for the Euthanasia of Animals,* 2013 ed. [online] Available at: https://www.avma.org/KB/Policies/Documents/euthanasia.pdf [Accessed 4 July 2017].

Lees, J.S., E.S. Sena, K.J. Egan, A. Antonic, S.A. Koblar, D.W. Howells and M.R. Macleod (2012). Stem Cell-based Therapy for Experimental Stroke: A Systematic Review and Meta-analysis. *International Journal of Stroke*, 7(7), pp. 582–588.

Leist, M. and T. Hartung (2013). Inflammatory Findings On Species Extrapolations: Humans Are Definitely No 70 kg Mice. *Archives of Toxicology*, 87(4), pp. 563–567.

Lewis, M.H., M.F. Presti, J.B. Lewis and L.A. Turner (2006). The neurobiology of stereotypy I. Environmental complexity. In: G.J. Mason and J. Rushen, eds., *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare,* 2nd ed. Oxfordshire: CAB International, pp. 190–226.

Li, G., Y. Gan, Y. Fan, Y. Wu, H. Lin, Y. Song, X. Cai, X. Yu, W. Pan, M. Yao, J. Gu and H. Tu (2015). Enriched Environment Inhibits Mouse Pancreatic Cancer Growth and Down-regulates the Expression of Mitochondria-related Genes in Cancer Cells. *Scientific Reports*, 5, p. 7856. [online] Avaiable at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4297951/ [Accessed 1 July 2017].

Lilley, E., R. Armstrong, N. Clark, P. Gray, P. Hawkins, K. Mason, N. López-Salesansky, A. Stark, S. Jackson, C. Thiemermann and M. Nandi (2015). Refinement of Animal Models of Sepsis and Septic Shock. *Shock*, 43(4), pp. 304–316.

Lidster, K., J.G. Jefferys, I. Blümcke, V. Crunelli, P.A. Flecknell, B.G. Frenguelli, W.P. Gray, R. Kaminski, A. Pitkänen, I. Ragan and M. Shah (2016). Opportunities for improving animal welfare in rodent models of epilepsy and seizures. *Journal of Neuroscience Methods*, 260, pp. 2–25.

Low, P. (2012). *Cambridge Declaration on Consciousness.* [online] Avaiable at: http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf [Accessed 1 July 2017].

Lush Prize (n.d.). *About the Lush Prize*. [online] Available at: http://lushprize.org/awards/ [Accessed 11 July 2018].

Macleod, M.R. (2011). Why Animal Research Needs To Improve: Many of The Studies That Use Animals To Model Human Diseases Are Too Small and Too Prone To Bias To Be Trusted. *Nature*, 477(7366), pp. 511–512.

Macleod, M.R., T. O'Collins, D.W. Howells and G.A. Donnan (2004). Pooling of Animal Experimental Data Reveals Influence Of Study Design and Publication Bias. *Stroke*, 35(5), pp. 1203–1208.

Macleod, M.R., A. Lawson McLean, A. Kyriakopoulou, S. Serghiou, A. de Wilde, N. Sherratt, T. Hirst, R. Hemblade, Z. Bahor, C. Nunes-Fonseca, A. Potluru, A. Thomson,

J. Baginskitae, K. Egan, H. Vesterinen, G. Currie, L. Churilov, D. Howells and E. Sena (2015). Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. *PLoS Biology*, 13(10), p. e1002273. [online] Available at: https://core.ac.uk/download/pdf/45473665.pdf [Accessed 11 July 2017].

Macleod, M.R., H.B. van der Worp, E.S. Sena, D.W. Howells, U. Dirnagl and G.A. Donnan (2008). Evidence for the Efficacy of NXY-059 in Experimental Focal Cerebral Ischaemia Is Confounded by Study Quality. *Stroke*, 39(10), pp. 2824–2829.

Mahoney, C.J. (1992). Some Thoughts on Psychological Enrichment. *Lab Animal,* 21(5), pp. 27–37.

Mak, I.W., N. Evaniew and M. Ghert (2014). Lost in Translation: Animal Models and Clinical Trials in Cancer Treatment. *American Journal of Translational Research*, 6(2), pp. 114–118.

Makowska, I.J. and D.M. Weary (2016a). Differences in Anticipatory Behaviour Between Rats (Rattus Norvegicus) Housed In Standard Versus Semi-naturalistic Laboratory Environments. *PLoS One*, 11(1), p. e0147595. [online] Available at: https://doi.org/10.1371/journal.pone.0147595 [Accessed 8 July 2017].

Makowska, I.J. and D.M. Weary (2016b). The Importance of Burrowing, Climbing, and Standing Upright for Laboratory Rats. *Royal Society Open Science*, *3*(6), p. 160136. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4929907/ [Accessed 8 July 2017].

Marks, J. (2012). Accept No Substitutes: The Ethics of Alternatives. *Hastings Center Report*, 42(s1).

Martin, B., S. Ji, S. Maudsley and M.P. Mattson (2010). "Control" Laboratory Rodents Are Metabolically Morbid: Why It Matters. *Proceedings of the National Academy of Sciences*, 107(14), pp. 6127–6133.

Martire, V.L., A. Silvani, S. Bastianini, C. Berteotti and G. Zoccoli (2012). Effects of Ambient Temperature on Sleep and Cardiovascular Regulation in Mice: The Role of Hypocretin/Orexin Neurons. *PLoS One*, 7(10), p. e47032. [online] Available at: https://doi.org/10.1371/journal.pone.0047032 [Accessed on 23 February 2017].

Mason, G.J. (2006). Stereotypic Behaviour in Captive Animals: Fundamentals and Implications for Welfare and Beyond. In: G. Mason and J. Rushen, eds., *Stereotypic Animal Behaviour: Fundamentals and Applications to Welfare*, 2nd ed. Oxfordshire: CAB International, pp. 325–356.

Maurin, J. (2012). Blutige Bisswunden, hungernde Mäuse. *TAZ.* [online] Available at: http://www.taz.de/!5095484/ [Accessed on 23 February 2017].

McManus, R. (2013). Ex-Director Zerhouni Surveys Value of NIH Research. *NIH Record*, LXV(13). [online] Available at: https://nihrecord.nih.gov/newsletters/2013/06_21_2013/story1.htm [Accessed on 23 February 2017].

McNutt, M., (2014). Journals Unite for Reproducibility. *Science*, 346(6210), p. 679.

Meagher, R.K. and G.J. Mason (2012). Environmental Enrichment Reduces Signs of Boredom in Caged Mink. *PLoS One*, 7(11), p. e49180. [online] Available at: https://doi.org/10.1371/journal.pone.0049180 [Accessed on 23 February 2017].

Meijer, M.K., B.M. Spruijt, L.F.M. Van Zutphen and V. Baumans (2006). Effect of Restraint and Injection Methods on Heart Rate and Body Temperature in Mice. *Laboratory animals*, 40(4), pp. 382–391.

Mergenthaler, P. and A. Meisel (2012). Do Stroke Models Model Stroke?. *Disease Models & Mechanisms*, 5(6), pp. 718–725.

Messmer, M.N., K.M. Kokolus, J.W.L. Eng, S.I. Abrams and E.A. Repasky (2014). Mild Cold-stress Depresses Immune Responses: Implications for Cancer Models Involving Laboratory Mice. *Bioessays*, 36(9), pp. 884–891.

Mestas, J. and C. Hughes (2004). Of Mice and Not Men: Differences Between Mouse and Human Immunology. *The Journal of Immunology*, 172(5), pp. 2731–2738.

Moher, D., A. Liberati, J. Tetzlaff, D.G. Altman, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6(7), p. e1000097. [online] Available at: https://doi.org/10.1371/journal.pmed.1000097 [Accessed 12 January 2017].

Morton, D. (2000). A Systematic Approach for Establishing Humane Endpoints. *ILAR Journal*, 41(2), pp. 80–86.

Mullane, K. and M. Williams (2014). Animal Models of Asthma: Reprise or Reboot?. *Biochemical Pharmacology*, 87(1), pp. 131–139.

National Animal Ethics Advisory Committee (2002). *A Culture of Care: A Guide for People Working With Animals in Research, Testing and Teaching.* Wellington, New Zealand. [online] Available at: https://anzccart.org.nz/app/uploads/2017/03/culture-of-care.pdf [Accessed 10 June 2017].

National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, D.C.: National Academies Press.

National Toxicology Program (2004). *A National Toxicology Program to the 21st Centruy. A Roadmap for the Future.* [online] Available at: https://ntp.niehs.nih.gov/ntp/about_ntp/ntpvision/ntproadmap_508.pdf [Accessed 11 January 2017].

NCad (n.d.). *NCad opinion Transition to non-animal research.* [online] Available at: https://www.ncadierproevenbeleid.nl/documenten/rapport/2016/12/15/ncad-opinion-transition-to-non-animal-research [Accessed 11 July 2017].

NCATS (National Center for Advancing Translational Sciences) (2017). *Tissue Chips for Disease Modeling and Efficacy Testing.* [online] Available at: https://ncats.nih.gov/tissuechip/projects/modeling [Accessed 14 July 2017].

NAVS (2018). *Fund Smarter Science*. [online] Available at: https://www.navs.org/what-we-do/fund-smarter-science/#.W1UAPi3Myu5 [Accessed 10 July 2018].

NC3Rs (n.d. a). *Funding Scheme Priority Areas*. [online] Available at: https://www.nc3rs.org.uk/funding-scheme-priority-areas#historic [Accessed 16 July 2018].

NC3Rs (n.d. b). *Experimental Design*. [online] Available at: https://nc3rs.org.uk/experimental-design [Accessed 14 July 2017].

NC3Rs (n.d. c). *Research Hubs*. [online] Available at: https://nc3rs.org.uk/resource-hubs [Accessed 14 July 2017].

NC3Rs (n.d. d). *Our funding*. [online] Available at: https://www.nc3rs.org.uk/about-us/funders [Accessed 16 July 2018].

Nerem, R.M., M.J. Levensque and J.F. Cornhill (1980). Social Environment as a Factor of Diet Induced Atherosclerosis. *Science,* 208, pp. 1475–1476.

NIH (2019). *Budget.* [online] Available at: https://www.nih.gov/about-nih/what-we-do/budget [Accessed 14 February 2019].

Nithianantharajah, J. and A. Hannan (2006). Enriched Environments, Experience-dependent Plasticity and Disorders of the Nervous System. *Nature Reviews Neuroscience*, 7(9), pp. 697–709.

NHMRC (National Health and Medical Research Council) (2013). *Australian code for the care and use of animals for scientific purposes.* 8th ed. [online] Available at: https://www.nhmrc.gov.au/guidelines-publications/ea28 [Accessed 11 July 2017].

Noor, F. (2019). The changing paradigm in preclinical toxicology: *in vitro* and *in silico* methods in liver toxicity evaluations. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Norecopa (2016a). *Culture of care*. [online] Available at: https://norecopa.no/more-resources/culture-of-care [Accessed 11 August 2018].

Norecopa (2016b). *Design and reporting of animal experiments*. [online] Available at: https://norecopa.no/more-resources/experimental-design-and-reporting [Accessed 12 August 2018].

Norecopa (2017). *Systematic Review Centre for Laboratory Animal Experimentation* (*SYRCLE*). [online] Available at: https://norecopa.no/3r-guide/systematic-review-centre-for-laboratory-animal-experimentation-syrcle [Accessed 12 August 2018].

O'Collins, V.E., M.R. Macleod, G.A. Donnan, L.L. Horky, B.H. van der Worp and D.W. Howells (2006). 1,026 Experimental Treatments in Acute Stroke. *Annals of Neurology*, 59(3), pp. 467–477.

Olson, H., G. Betton, D. Robinson, K. Thomas, A. Monro, G. Kolaja, P. Lilly, J. Sanders, G. Sipes, W. Bracken, M. Dorato, K. Van Deun, P. Smith, B. Berger and A. Heller (2000). Concordance of the Toxicity of Pharmaceuticals in Humans and in Animals. *Regulatory Toxicology and Pharmacology*, 32(1), pp. 56–67. [online] Available at: https://www.gwern.net/docs/statistics/meta-analysis/2000-olson.pdf [Accessed 12 August 2017].

Olsson, I.A.S., C.M. Nevison, E.G. Patterson-Kane, C.M. Sherwin, H.A. van de Weerd and H. Würbel (2003). Understanding Behaviour: The Relevance of Ethological Approaches in Laboratory Animal Science. *Applied Animal Behaviour Science*, 81(3), pp. 245–264.

Ormandy, E.H., C.A. Schuppli, and D.M. Weary (2009). Worldwide Trends in the Use of Animals in Research: The Contribution of Genetically-modified Animal Models. *Alternatives to Laboratory Animals*, 37, pp. 63–68.

Osborne, N., Avey, M.T., Anestidou, L., Ritskes-Hoitinga, M. and Griffin, G. (2018). Improving animal research reporting standards: HARRP, the first step of a unified approach by ICLAS to improve animal research reporting standards worldwide. EMBO Reports , 19(5), p. e46069. [online] Available at: http://embor.embopress.org/content/19/5/e46069 [Accessed 19 April 2018].

Pawlowski, J., D. Feinstein, M.L. Crandall and S. Gala (2019). Modernizing biomedical training: replacing live animal laboratories with human simulation. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill, pp. 551–66.

Payne, K.J. and G.M. Crooks (2007). Immune-cell Lineage Commitment: Translation From Mice to Humans. *Immunity*, 26(6), pp. 674–677.

Percie du Sert, N., A. Alfieri, S.M. Allan, H.V. Carswell, G.A. Deuchar, T.D. Farr, P. Fleck-nell, L. Gallagher, C.L. Gibson, M.J. Haley and M.R. Macleod (2017). The IMPROVE Guidelines (Ischaemia Models: Procedural Refinements Of in Vivo Experiments). *Journal of Cerebral Blood Flow & Metabolism*, p. 0271678X17709185. [online] Available at: http://journals.sagepub.com/doi/pdf/10.1177/0271678X17709185 [Accessed 18 August 2017].

Perel, P., I. Roberts, E. Sena, P. Wheble, C. Briscoe, P. Sandercock, M. Macleod, L.E. Mignini, P. Jayaram and K.S. Khan (2007). Comparison of Treatment Effects Between Animal Experiments and Clinical Trials: Systematic Review. *British Medical Journal*, 334, pp. 197–206. [online] Available at: https://www.bmj.com/content/334/7586/197 [Accessed 12 July 2017].

PETA International Science Consortium (PISC) (n.d. a). *Awards of VITROCELL In Vitro Inhalation Exposure Systems.* [online] Available at: https://www.piscltd.org.uk/vitrocell-prize/ [Accessed 10 August 2017].

PETA International Science Consortium (PISC) (n.d. b). *Funding to Advance the Development and Use of Non-Animal Methods.* [online] Available at: https://www.piscltd.org.uk/funding/ [Accessed 10 August 2018].

Pew Research Center (2015). *Public and Scientists' Views on Science and Society: Use of Animals in Scientific Research.* [online] Available at: http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/pi_2015-01-29_science-and-society-03-05/ [Accessed 10 July 2017].

Pew Research Center (2018). Americans are divided over the use of animals in scientific research. [online] Available at: http://www.pewresearch.org/fact-tank/2018/08/16/americans-are-divided-over-the-use-of-animals-in-scientific-research/. [Accessed 10 July 2017].

Pharmaceutical Research and Manufacturers of America (2015). *Biopharmaceutical research industry profile.* Washington, DC: PhRMA. [online] Available at: http://

phrma-docs.phrma.org/sites/default/files/pdf/2015_phrma_profile.pdf [Accessed 10 July 2017].

Pharmaceutical Research and Manufacturers of America (2016). Biopharmaceutical research industry profile. Washington, DC: PhRMA. [online] Available at: http://phrma-docs.phrma.org/sites/default/files/pdf/biopharmaceutical-industry-profile.pdf [Accessed 10 December 2017].

Pippin, J.J. (2012). Animal Research in Medical Sciences: Seeking a Convergence of Science, Medicine, and Animal Law. *South Texas Law Review,* 54, pp. 469–511.

Pippin, J.J., S.E. Cavanaugh and F. Pistollato (2019). Animal Research for Alzheimer Disease: Failures of Science and Ethics. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Pistollato, F., E.L. Ohayon, A. Lam, G.R. Langley, T.J. Novak, D. Pamies, G. Perry, E. Trushina, R.S.B. Williams, A.E. Roher, T. Hartung, S. Harnard, N. Barnard, M.C. Morris, M. Lai, R. Merkley and P.C. Chandrasekera (2016). Alzheimer Disease research in the 21st century: past and current failures, new perspectives and funding priorities. *Oncotarget*, 7(26), pp. 38999–39016. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5129909/ [Accessed 10 July 2017].

Poole, T. (1997). Happy Animals Make Good Science. *Laboratory Animals*, 31(2), pp. 116–124.

Pound, P. and M.B. Bracken (2014). Is Animal Research Sufficiently Evidence Based To Be a Cornerstone of Biomedical Research?. *British Medical Journal,* 348, p. g3387.

Pound, P., S. Ebrahim, P. Sandercock, M.B. Bracken and I. Roberts (2004). Where Is the Evidence That Animal Research Benefits Humans?. *British Medical Journal*, 328(7438), pp. 514–517.

Pound, P. and C.J. Nicol (2018). Retrospective Harm-benefit Analysis of Preclinical Animal Research for Six Treatment Interventions. *PLoS One,* 13(3), p. e0193758. [online] Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193758 [Accessed 28 March 2018].

Pound, P. and Ritskes-Hoitinga, M. (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(1), p. 304. [online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6223056/ [Accessed 11 November 2018].

Preclinical Trials (n.d.). *International Register for Preclincial Protocols.* [online] Available at: https://www.preclinicaltrials.eu [Accessed 10 December 2017].

Prinz, F., T. Schlange and K. Asadullah (2011). Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?. *Nature Reviews Drug Discovery*, 10(9), pp. 712–712.

Ram, R. (2019). Extrapolation of animal research data to humans: an analysis of the evidence. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Ramirez, T., S. Beken, M. Chlebus, G. Ellis, C. Griesinger, S. De Jonghe, I. Manou, A. Mehling, K. Reisinger, L. Rossi, J. van der Laan, R. Weissenhorn and U. Sauer (2015). Knowledge sharing to facilitate regulatory decision-making in regard to alternatives to animal testing: Report of an EPAA workshop. *Regulatory Toxicology and Pharmacology*, 73(1), pp. 210–226.

Redmond, C. (2019). When is an alternative not an alternative? Supporting progress for absolute replacement of animals in science. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22. Leiden: Brill.

Renner, M., M.A. Lancaster, S. Bian, H. Choi, T. Ku, A. Peer, K. Chung and J.A. Knoblich (2017). Self-organized Developmental Patterning and Differentiation in Cerebral Organoids. *The EMBO Journal*, p. e201694700. [online] Available at: http://emboj.embopress.org/content/36/10/1316.long [Accessed 10 May 2017].

Reichlin, T.S., L. Vogt and H. Würbel (2016). The Researchers' View of Scientific Rigor—Survey on the Conduct and Reporting of In Vivo Research. *PLoS One*, 11(12), p. e0165999. [online] Available at: https://doi.org/10.1371/journal.pone.0165999 [Accessed 10 January 2017].

Reinhardt, V. (2003). Compassion for Animals in the Laboratory: Impairment or Refinement of Research Methodology?. *Journal of Applied Animal Welfare Science*, 6(2), pp. 123–130.

Richardson, C.A. and P.A. Flecknell (2005). Anaesthesia and Postoperative Analgesia Following Experimental Surgery in Laboratory Rodents: Are We Making Progress?. *Alternatives to Laboratory Animals*, 33, pp. 119–127.

Richter, S.H. (2017). Systematic Heterogenization for Better Reproducibility in Animal Experimentation. *Lab Animal*, 46(9), pp. 343–349.

Richter, S.H., J.P. Garner and H. Würbel (2009). Environmental Standardization: Cure or Cause of Poor Reproducibility in Animal Experiments?. *Nature Methods*, 6(4), pp. 257–261.

Richter, S.H., J.P. Garner, C. Auer, J. Kunert and H. Würbel (2010). Systematic Variation Improves Reproducibility of Animal Experiments. *Nature Methods*, 7(3), pp. 167–168.

Ritskes-Hoitinga, M. (2016). Systematic Reviews of Animal Studies Equal the Implementation of the 3Rs. In: *20th European Congress on Alternatives to Animal Testing, 17th Annual Congress of EUSAAT*. Linz. [online] Availabe at: https://norecopa.no/media/7597/systematic-reviews.pdf [Accessed 15 December 2017].

Ritskes-Hoitinga, M. and K. Wever (2018). Improving the Conduct, Reporting, and Appraisal of Animal Research. Editorial. *British Medical Journal*, 360, p. j4935.

Rittirsch, D., L. Hoesel and P. Ward (2007). The Disconnect Between Animal Models of Sepsis and Human Sepsis. *Journal of Leukocyte Biology*, 81(1), pp. 137–143.

Rosenzweig, M.R. (1966). Environmental Complexity, Cerebral Change, and Behavior. *American Psychologist,* 21, pp. 321–332.

Rosenzweig, M.R. and E.L. Bennett (1969). Effects of Differential Environments on Brain Weights and Enzyme Activities in Gerbils, Rats, and Mice. *Developmental Psychobiology*, 2(2), pp. 87–95.

Rosenzweig, M.R., D. Krech, E.L. Bennett and J.F. Zolman (1962). Variation in Environmental Complexity and Brain Measures. *Journal of Comparative and Physiological Psychology*, 55(6), p. 1092.

Røttingen, J.A., S. Regmi, M. Eide, A.J. Young, R.F. Viergever, C. Årdal, J. Guzman, D. Edwards, S.A. Matlin and R.F. Terry (2013). "Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory?" *The Lancet*, 382 (9900), pp. 1286–1307.

Rovida, C., S. Asakura, M. Daneshian, H. Hofman-Huether, M. Leist, L. Meunier, D. Reif, A Rossi, M. Schmutz, J. Valentin, J. Zurlo and T. Hartung (2015). Toxicity Testing in the 21st Century Beyond Environmental Chemicals. *Alternatives to Animal Experimentation*, 32(3), pp. 171–181.

Russell, W.M.S. and R.L. Burch (1959). *The Principles of Humane Experimental Technique*. Potters Bar, Hertfordshire, England: Universities Federation for Animal Welfare. [online] Available at: http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc [Accessed 12 May 2017].

Salzberg, S.L., O. White, J. Peterson and J.A. Eisen (2001). Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss?. *Science*, 292 (5523), pp. 1903–1906.

Sandercock, P. and I. Roberts (2002b). Systematic Reviews of Animal Experiments. *The Lancet*, 360, p. 586.

Scannell, J.W. and Bosley, J. (2016). When Quality Beats Quantity: Decision Theory, Drug Discovery, and the Reproducibility Crisis. *PLoS One*, 11(2), p.e0147215. [online] Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147215 [Accessed 14 September 2017].

Scherer, R.W., Meerpohl, J.J., Pfeifer, N., Schmucker, C., Schwarzer, G. and von Elm, E. (2018). Full publication of results initially presented in abstracts. Cochrane Database of Systematic Reviews 2018, Issue 11. Art. No.: MR000005. DOI: 10.1002/14651858.MR000005.pub4.

Schmidt, D.A., M.R. Ellersieck, M.R. Cranfield and W.B. Karesh (2006). Cholesterol Values in Free-ranging Gorillas (Gorilla Gorilla Gorilla and Gorilla Beringei) and Bornean Orangutans (Pongo Pygmaeus). *Journal of Zoo and Wildlife Medicine*, 37(3), pp. 292–300.

Schuppli, C., A. Walterhouse, V. Chew, N. Hammound, L. Kolody, B. Tan, J. Makowska, S. Mcnamara, J. Sato-Reinhold, V. Wong and D. Weary (2017). A Better Life for Research Animals by Fostering a Culture of Compassion Amongst Researchers. In: *10th World Congress on Alternatives and Animal Use in the Life Sciences* (*WC10*). Seattle: Alternatives to Animal Experimentation Proceedings, 6(1), p. 13. [online] Available at: http://www.altex.ch/resources/WC10_entire_issue1.pdf [Accessed 12 September 2017].

Sena, E.S., H.B. van der Worp, P.M. Bath, D.W. Howells and M.R. Macleod (2010). Publication Bias in Reports of Animal Stroke Studies Leads To Major Overstatement of Efficacy. *PLoS Biology*, 8(3), p. e1000344. [online] Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000344 [Accessed 12 August 2017].

Seok, J., H.S. Warren, A.G. Cuenca, M.N. Mindrinos, H.V. Baker, W. Xu, D.R. Richards, G.P. McDonald-Smith, H. Gao, L. Hennessy and C.C. Finnerty (2013). Genomic Responses in Mouse Models Poorly Mimic Human Inflammatory Diseases. *Proceedings of the National Academy of Sciences*, 110(9), pp. 3507–3512.

Serrat, M.A., D. King and C.O. Lovejoy (2008). Temperature Regulates Limb Length in Homeotherms By Directly Modulating Cartilage Growth. *Proceedings of the National Academy of Sciences*, 105 (49), pp. 19348–19353.

Shay, T., V. Jojic, O. Zuk, K. Rothamel, D. Puyraimond-Zemmour, T. Feng, E. Wakamatsu, C. Benoist, D. Koller and A. Regev (2013). Conservation and Divergence in the Transcriptional Programs of the Human and Mouse Immune Systems. *Proceedings of the National Academy of Sciences*, 110(8), pp. 2946–2951.

Sherwin, C.M. (2004). The influences of Standard Laboratory Cages on Rodents and the Validity of Research Data. *Animal Welfare*, 13(1), pp. 9–15.

Sherwin, C.M., E. Haug, N. Terkelsen and M. Vadgama (2004). Studies on the Motivation for Burrowing by Laboratory Mice. *Applied Animal Behaviour Science*, 88(3), pp. 343–358.

Shuaib, A., K. Lees, P. Lyden, J. Grotta, A. Davalos, S. Davis, H. Diener, T. Ashwood, W. Wasiewski and U. Emeribe (2007). NXY-059 for the Treatment of Acute Ischemic Stroke. *New England Journal of Medicine*, 357(6), pp. 562–571.

Simmons, J.P., L.D. Nelson and U. Simonsohn (2011). False-positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22 (11), pp. 1359–1366.

Sloff, M., R. de Vries, P. Geutjes, J. IntHout, M. Ritskes-Hoitinga, E. Oosterwijk and W. Feitz (2014). Tissue Engineering in Animal Models for Urinary Diversion: A Systematic Review. *PLoS one*, 9(6), p. e98734. [online] Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098734 [Accessed 30 July 2017].

Smith, A.J., R.E. Clutton, E. Lilley, K.E.A. Hansen and T. Brattelid (2017). PREPARE: Guidelines for Planning Animal Research and Testing. *Laboratory Animals*, pp. 135–141. [online] Available at: http://journals.sagepub.com/doi/pdf/10.1177/0023677217724823 [Accessed 5 August 2017].

Smith, R. (2014). Medical Research—Still a Scandal. [Blog] *British Medical Journal Blog*. Available at: https://blogs.bmj.com/bmj/2014/01/31/richard-smith-medical-research-still-a-scandal/ [Accessed 30 August 2017].

Sorge, R., L. Martin, K. Isbester, S. Sotocinal, S. Rosen, A. Tuttle, J. Wieskopf, E. Acland, A. Dokova, B. Kadoura, P. Leger, J. Mapplebeck, M. McPhail, A. Delaney, G. Wigerblad, A. Schumann, T. Quinn, J. Frasnelli, C. Svensson, W. Sternberg and J. Mogil (2014). Olfactory Exposure to Males, Including Men, Causes Stress and Related Analgesia in Rodents. *Nature Methods*, 11(6), pp. 629–632.

Špinka, M. and F. Wemelsfelder (2011). Environmental challenge and animal agency. In: M.C. Appleby, J.A. Mench, I.A.S. Olsson and B.O. Hughes, eds., *Animal Welfare,* 2nd ed. Wallingford: CABI International, pp. 27–43.

Stokes, E., P. Flecknell and C. Richardson (2009). Reported Analgesic and Anaesthetic Administration To Rodents Undergoing Experimental Surgical Procedures. *Laboratory Animals*, 43(2), pp. 149–154.

Systematic Review Center for Laboratory Animal Experimentation (SYRCLE) (n.d. a). *Systematic reviews of animal studies*. [video] Available at: https://www.radboudumc.nl/en/research/radboud-technology-centers/animal-research-facility/systematic-review-center-for-laboratory-animal-experimentation [Accessed 12 July 2017].

SYRCLE (n.d. b). *Tools and guidelines*. [online] Available at: https://www.radboudumc.nl/en/research/radboud-technology-centers/animal-research-facility/systematic-review-center-for-laboratory-animal-experimentation [Accessed 12 July 2017].

SYRCLE (n.d. c). Protocols. Turner et al. (2017). The use of carbon dioxide as a method for euthanasia of laboratory mice and rats – a systematic review. [online] Available at: https://issuu.com/radboudumc/docs/the_use_of_carbon_dioxide_as_a_meth?e=28355229/48248733 [Accessed 12 January 2018].

Taylor, K. and L. Rego (2016). EU statistics on Animal Experiments for 2014. *Alternatives to Animal Experimentation Proceedings*, 33(4), pp. 465–468.

Taylor, K., N. Gordon, G. Langley and W. Higgins (2008). Estimates for Worldwide Laboratory Animal Use in 2005. *Alternatives to Laboratory Animals*, 36(3), pp. 327–342.

Tsai, P.P., U. Pachowsky, H.D. Stelzer and H. Hackbarth (2002). Impact of Environmental Enrichment in Mice. 1: Effect of Housing Conditions on Body Weight, Organ Weights and Haematology in Different Strains. *Laboratory Animals*, 36(4), pp. 411–419.

Tsai, P.P., H.D. Stelzer, H.J. Hedrich and H. Hackbarth (2003). Are the Effects of Different Enrichment Designs on the Physiology and Behaviour of DBA/2 Mice Consistent?. *Laboratory Animals*, 37(4), pp. 314–327.

Tsai, P.P., H.D. Stelzer, A. Schraepler and H. Hackbarth (2006). Importance and Effects of Enrichment on Physiology, Behaviour and Breeding Performance in Mice. *Alternatives to Animal Experimentation*, 23, pp. 96–98.

Tsilidis, K., O. Panagiotou, E. Sena, E. Aretouli, E. Evangelou, D. Howells, R. Salman, M. Macleod and J.P.A. Ioannidis (2013). Evaluation of Excess Significance Bias in Animal Studies of Neurological Diseases. *PLoS Biology*, 11(7), p. e1001609. [online] Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001609 [Accessed 12 January 2018].

Tsukamoto, T. (2016) Animal Disease Models for Drug Screening: The Elephant in the Room? *Drug Discovery Today*, 21, pp. 529–530.

Turner, J.G., J.L. Parrish, L.F. Hughes, L.A. Toth and D.M. Caspary (2005). Hearing in Laboratory Animals: Strain Differences and Nonauditory Effects of Noise. *Comparative Medicine*, 55(1), pp. 12–23.

UC San Diego (2018). *Biomedical Research using Non Animal Models.* [online] Available at: https://extension.ucsd.edu/courses-and-programs/biomedical-research-using-non-animal-models [Accessed 12 July 2018].

Uhlig, C., H. Krause, T. Koch, M.G. de Abreu and P.M. Spieth (2015). Anesthesia and Monitoring in Small Laboratory Mammals Used in Anesthesiology, Respiratory and Critical Care Research: A Systematic Review on the Current Reporting in top-10 Impact Factor Ranked Journals. *PLoS One,* 10(8), p. e0134205. [online] Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134205 [Accessed 18 July 2017].

van de Weerd, H.A., Aarsen, E.L., Mulder, A., Kruitwagen, C.L., Hendriksen, C.F. and Baumans, V. (2002). Effects of environmental enrichment for mice: variation in experimental results. *Journal of Applied Animal Welfare Science*, 5(2), pp. 87–109.

van der Poll, T. (2012). Experimental Human Sepsis Models. *Drug Discovery Today: Disease Models*, 9(1), pp. e3–e9.

van der Worp, H.B., D.W. Howells, E.S. Sena, M.J. Porritt, S. Rewell, V. O and M.R. Macleod (2010). Can Animal Models of Disease Reliably Inform Human Studies? *PLoS Medicine*, 7, pp. e1000245. [online] Available at: http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000245 [Accessed 15 August 2017].

van Luijk, J. (2016). Improving Animal Research Using a Science Driven Approach: Sytematic Reviews of Animal Studies. In: *20th European Congress on Alternatives to Animal Testing, 17th Annual Congress of EUSAAT*. Linz. [online] Availabe at: https://norecopa.no/media/7599/systematic-reviews-judith-van-luijk.pdf [Accessed 15 December 2017].

van Luijk, J., B. Bakker, M.M. Rovers, M. Ritskes-Hoitinga, R.B. de Vries and M. Leenaars (2014). Systematic Reviews of Animal Studies: Missing Link in Translational Research?. *PLoS One*, 9(3), p. e89981. [online] Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089981 [Accessed 15 August 2017].

van Luijk, J., Y. Cuijpers, L. van der Vaart, T.C. de Roo, M. Leenaars and M. Ritskes-Hoitinga (2013). Assessing the Application of the 3Rs: A Survey Among Animal Welfare Officers in the Netherlands. *Laboratory Animals*, 47(3), pp. 210–219.

van Luijk, J., M. Leenaars, A.M. van Dongen, L. van der Vaart and M. Ritskes-Hoitinga (2012). Outcomes of a Dutch Workshop on Improvements for the 3Rs in Daily Practice. *Alternatives to Animal Experimentation*, 29(4), p. 440–443. [online] Available at: https://www.altex.org/index.php/altex/article/view/439/447 [Accessed 15 August 2017].

van Praag, H., G. Kempermann and F.H. Gage (2000). Neural Consequences of Environmental Enrichment. *Nature Reviews Neuroscience*, 1(3), pp. 191–198.

vfa (Die forschenden Pharmaunternehmen) (2017). *Tierschutz. Preise des Landes Berlin für tierfreie Forschung zu Grippeviren und Schlaganfall verliehen.* [online] Available at: https://www.vfa.de/de/arzneimittel-forschung/forschungsstandort-deutschland/preise-des-landes-berlin-fuer-tierfreie-forschung-zu-grippeviren-und-schlaganfall-verliehen.html [Accessed 12 October 2017].

Voelkl, B., Vogt, L., Sena, E.S. and Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biology,* 16(2), p. e2003693.

Vogt, L., T.S. Reichlin, C. Nathues and H. Würbel (2016). Authorization of Animal Experiments Is Based on Confidence Rather Than Evidence of Scientific Rigor. *PLoS Biology, 14* (*12*), p. e2000598. [online] Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2000598 [Accessed 12 July 2017].

von Aulock, S. (2017). The 3Rs in Action, Session VIII-3: Establishing a Culture of Care Through Assessment, Transparency, and Communication. In: *10th World Congress on Alternatives and Animal Use in the Life Sciences* (*WC10*). Seattle: Alternatives to Animal Experimentation Proceedings, 6(1), p. 213. [online] Available at: http://www.altex.ch/resources/WC10_entire_issue1.pdf [Accessed 22 August 2017].

Waiblinger, E. and B. König (2004). Refinement of Gerbil Housing and Husbandry in the Laboratory. *Alternatives to Laboratory Animals*, 32 (1A), pp. 163–169.

Watts, N., M. Amann, S. Ayeb-Karlsson, K. Belesova, T. Bouley, M. Boykoff, P. Byass, W. Cai, D. Campbell-Lendrum, J. Chambers and P.M. Cox (2017). The Lancet Countdown on Health and Climate Change: From 25 Years of Inaction to a Global Transformation for Public Health. *The Lancet*, 391 (10120), pp. 581–630.

Weed, J. and J. Raber (2005). Balancing Animal Research With Animal Well-being: Establishment of Goals and Harmonization of Approaches. *ILAR Journal*, 46(2), pp. 118–128.

Wells, S. (2017). Pain Assessment and New Innovations. In: *10th World Congress on Alternatives and Animal Use in the Life Sciences* (*WC10*). Seattle: Alternatives to Animal Experimentation Proceedings, 6(1), p. 222. [online]Available at: http://www.altex.ch/resources/WC10_entire_issue1.pdf [Accessed 22 August 2017].

Wever, K.E., F.J. Geessink, M.A. Brouwer, A. Tillema and M. Ritskes-Hoitinga (2017). A systematic review of discomfort due to toe or ear clipping in laboratory rodents. *Laboratory Animals*, 51(6), pp. 583–600.

Wiedenmayer, C. (1997). Causation of the Ontogenetic Development of Stereotypic Digging in Gerbils. *Animal Behaviour*, 53(3), pp. 461–470.

Wolfensohn, S., P. Hawkins, E. Lilley, D. Anthony, C. Chambers, S. Lane, M. Lawton, H. Voipio and G. Woodhall (2013a). Reducing Suffering in Experimental Autoimmune Encephalomyelitis (EAE). *Journal of Pharmacological and Toxicological Methods*, 67(3), pp. 169–176.

Wolfensohn, S., P. Hawkins, E. Lilley, D. Anthony, C. Chambers, S. Lane, M. Lawton, S. Robinson, H. Voipio and G. Woodhall (2013b). Reducing Suffering in Animal Models and Procedures Involving Seizures, Convulsions and Epilepsy. *Journal of Pharmacological and Toxicological Methods*, 67(1), pp. 9–15.

Wolfer, D., O. Litvin, S. Morf, R. Nitsch, H. Lipp and H. Würbel (2004). Laboratory Animal Welfare: Cage Enrichment and Mouse Behaviour. *Nature*, 432(7019), pp. 821–822.

Wolfle, T.L. (1987). Control of Stress Using Non-drug Approaches. *Journal of the American Veterinary Medical Association,* 191, pp. 1219–1221.

World Organisation for Animal Health (OIE) (2018). *Terrestrial Animal Health Code, Ch. 7.8*. Paris: World Organisation for Animal Health, first adopted in 2010; most recent

update adopted in 2013. [online] Available at: http://www.oie.int/index.php?id=169&L=0&htmfile=chapitre_aw_research_education.htm [Accessed 12 May 2017].

Würbel, H. (2000). Behaviour and the Standardization Fallacy. *Nature Genetics,* 26, p. 263.

Würbel, H. (2001). Ideal homes? Housing Effects on Rodent Brain and Behaviour. *Trends in Neuroscience*. 24, pp. 207–211.

Würbel, H. (2007). Environmental Enrichment Does Not Disrupt Standardisation of Animal Experiments. *Alternatives to Animal Experimentation*, 24, pp. 70–73.

Würbel, H. (2016). Tierversuch und Irrtum. In: M. Fehlmann, M. Michel and R. Niederhauser, eds. *Tierisch! Das Tier und die Wissenschaft: Ein Streifzug durch die Disziplinen*, Reihe Zürcher Hochschulforum, 55, pp. 97–106.

Würbel, H. (2017). More Than 3Rs: The Importance of Scientific Validity for Harm-benefit Analysis of Animal Research. *Nature Lab Animal*, 46(4), pp. 164–166. [online] Available at: https://www.nature.com/articles/laban.1220.pdf [Accessed 12 June 2017].

Würbel, H. and J.P. Garner (2007). Refinement of Rodent Research Through Environmental Enrichment and Systematic Randomization. *National Centre for the Replacement, Refinement, and Reduction of Animals in Research* (*NC3Rs*). [online] Available at: https://www.nc3rs.org.uk/sites/default/files/documents/Refinementenvironmentalenrichmentandsystematicrandomization.pdf [Accessed 12 August 2017].

Würbel, H. and M. Stauffacher (1994). Standardhaltung für Labormäuse—Probleme und Lösungsansätze. *Tierlaboratorium*, 17, pp. 109–118.

Würbel, H. and M. Stauffacher (1996). Prevention of Stereotypy in Laboratory Mice: Effects on Stress Physiology and Behaviour. *Physiology & Behavior*, 59(6), pp. 1163–1170.

Würbel, H., M. Stauffacher and D. Holst (1996). Stereotypies in Laboratory Mice—Quantitative and Qualitative Description of the Ontogeny of 'Wire-gnawing'and 'Jumping'in Zur: ICR and Zur: ICR nu. *Ethology*, 102(3), pp. 371–385.

Zeeff, S.B., C. Kunne, G. Bouma, R.B. de Vries and A.A. Te Velde (2016). Actual usage and Quality of Experimental Colitis Models in Preclinical Efficacy Testing: A Scoping Review. *Inflammatory Bowel Diseases*, 22(6), pp. 1296–1305.

Zerhouni, E.A. (2014). Turning the Titanic. *Science Translational Medicine*, 6(221), p. 221ed2.

Zurlo, J. (2012). No Animals Harmed: Toward a Paradigm Shift in Toxicity Testing. *Hastings Center Report*, 42(s1). [online] Available at: http://animalresearch.thehastingscenter.org/report/no-animals-harmed-toward-a-paradigm-shift-in-toxicity-testing/ [Accessed 12 June 2017].

Zurlo, J., D. Rudacille and A.M. Goldberg (1996). The Three R's: The Way Forward. *Environmental Health Perspectives*, 104(8), pp. 878–880.

2.5 Study findings

Original research: overview
The primary research of this study on the proposed application of experimental Refinements in animal research protocols, found shortcomings in all areas assessed: anesthesia, peri- and postoperative analgesia, postoperative care and monitoring for adverse effects, humane endpoints, killing methods, and the researchers' severity estimates of their planned surgical procedures. The deficiencies in the application of Refinement methods can lead to avoidable pain and suffering for the animal subjects. According to Section 9, para. 2, number 3, *German Animal Welfare Act* of 1998, pain, suffering and harms are only allowed to be inflicted upon the research animals to the degree that is indispensable to reach the research goal; and investigators are responsible for using experimental methods, incl. Refinement methods, that are in accordance with current scientific knowledge (Section 9, para. 2, sentence 2).

Our study found that pain was often not alleviated or was unlikely to be alleviated effectively (see Paper 1, Herrmann and Flecknell, 2019). Following 30% of surgical procedures no analgesia was planned at all; in another 10% of procedures, pain would only be alleviated if it was observed. However, monitoring appeared to be insufficient to properly detect pain and suffering. Reasons for withholding pain relief, if included, implied a general paucity of monitoring, a lack of pain assessment and a possible lack of pain recognition skills. The study also found that concerns about the negative effects of analgesics on study outcomes were seemingly greater than the concerns about untreated pain.

Considering the severity of proposed surgical interventions (e.g., 17% of all surgical interventions on mice were thoracotomies), it was concerning that postsurgical multimodal pain management was almost never used. Preventive analgesia was often not provided when, for example, an anesthetic was used which had no analgesic properties (such as pentobarbital or isoflurane); this was the case in a quarter of these procedures when isoflurane was used, and in another 15% of cases with isoflurane, it was given too late, and thus, the animals would regain consciousness without analgesic coverage. In addition, the use of local anesthetics for severe surgeries such as thoracotomies was low (10% of thoracotomies). Furthermore, the study's findings showed that the severity of pain, distress, suffering and lasting harm that was to be inflicted upon the animals was frequently underestimated by the researchers; planned health monitoring appeared insufficient to ensure the welfare of animals in the studies (see Paper 2, Herrmann and Flecknell, 2018a).

In those cases, where the researchers included humane endpoints in their application, these were mostly not clearly defined and/or were at a late stage in the animal's deterioration and thus, are likely to have caused suffering that could have otherwise been avoided (Paper 3, Herrmann and Flecknell, 2018b). The choice of killing method could be improved in many applications, thereby reducing suffering. For instance, carbon dioxide (CO_2) was widely used in our study sample, either to anesthetize animals prior to killing by another method, or it was used as the single means to kill the animals. Because the appropriateness of carbon dioxide has been under scrutiny at least since the 1980s (Freed, 1983), and because of its large-scale use, a multitude of research studies investigated CO_2 and confirmed its aversiveness in several species including humans (for an overview, see e.g., EFSA AHAW Panel, 2005 and Hawkins *et al.*, 2006). It causes fear and anxiety in animals, e.g. in mice (Ziemann *et al.*, 2009), rats (Johnson *et al.*, 2011) and humans (Finlay *et al.* 2009). Furthermore, carbon dioxide provokes dyspnea (Banzett and Moosavi, 2001; Banzett *et al.*, 2008) and distress (Brofman *et al.*, 1990; Marquardt *et al.*, 2018). CO_2 inhalation at higher concentrations causes pain in animals including humans (Anton

et al., 1992; Danneman *et al.*, 1997; Peppel and Anton, 1993) (for a detailed discussion, see Paper 3).

At the time the assessed proposals were submitted to the German competent authorities, two guidance documents on how to estimate and classify the severity of various experimental procedures and interventions were available in German: one from the Swiss competent authority (Bundesamt für Veterinärwesen, 1995), which is legally binding in Switzerland, and one guide composed by the Group of Berlin Animal Welfare Officers (Arbeitskreis Berliner Tierschutzbeauftragter, 2010), which is based on the Swiss guide (Bundesamt für Veterinärwesen, 1995). It is unknown how many of the investigators in our sample used these guidance documents, when estimating the impacts on the animals by their surgical interventions. 59% (377/642) of the procedures rated in the applications deviated from estimates we made using expert guidance documents, and for 375 of the 377 procedures, the investigators' severity classification was lower. If research workers are adequately trained in animal behavior and in the recognition of species-specific distress, pain and suffering, they should have the expertise to assess the actual severity correctly. However, the frequent underestimation of anticipated pain and suffering found in this study (Paper 2) is in line with the scheduled underuse of proper anesthesia and analgesia regimens (Paper 1) and the shortcomings in regards to planned monitoring practices (Papers 2 to 3), the application of humane endpoints, and the choice of killing methods (Paper 3). All of this strongly indicates that discomfort and pain were not observed, not recognized and /or not taken seriously by many animal researchers.

Literature review

The brief overview (Paper 4; Herrmann, 2019) of information concerning Refinement research and its practical application in animal housing, husbandry, care, experimental procedures as well as in planning, conducting and reporting of animal studies, indicated shortcomings in the translation of these findings into practice. Paper 4 also uncovered that the goal of moving towards the replacement of (live) animals in all areas of scientific use, declared for example, by parts of the scientific community (see The Three Rs Declaration of Bologna (ATLA Staff Writer, 2000)) and by EU politicians (see Directive 2010/63/EU on the protection of animals used for scientific purposes (European Parliament, 2010)), is currently not realized to an extent that would help to reduce the growing numbers of animals used (European Commission, 2013; Taylor and Rego, 2016). In regulatory testing (Basketter *et al.*, 2012) and in education and training (Pawlowski *et al.*, 2019), some efforts have been made to find non-animal methods. However, in basic and applied research, for which the most animals are used (Daneshian *et al.*, 2015; European Commission, 2011) and the majority of procedures are conducted (Taylor and Rego Alvarez, 2019), animals still represent the default model. The underlying problems are multifactorial, with main issues probably being a lack of appreciation and knowledge about not only Refinement (as we identified in Papers 1 to 3) but in all the 3Rs (see Paper 4) as well as a lack of knowledge about the importance of searching for models that reflect the biology and the pathobiology of humans, rather than that of non-human animals (Archibald *et al.*, 2018; Herrmann *et al.*, 2019; Ritskes-Hoitinga and Pound, 2018); the lack of legal enforcement of the 3Rs in their hierarchical order where Replacement comes first and Refinement last (Russell and Burch, 1959; European Parliament, 2010); and the lack of incentives provided to scientists willing to change the status quo.

Animal welfare provisions not fulfilled

Legal requirements to safeguard animals used in science were in place well before 2010. It has been an obligation to reduce to an absolute minimum any pain, suffering and harms inflicted upon research animals since 1972 (Section 9, para. 1, sentence 1, No.1 *German Animal Welfare Act* as amended on 24 July 1972). In addition, since 1986, it has been illegal to cause animals

pain, suffering and/or harm in order to reduce workload, reduce the time needed to properly care for the animals or to cut costs (Section 9, para. 2, No. 3, *German Animal Welfare Act* as amended in 1986). The harm-benefit analysis (HBA) has been a cornerstone of the German animal research regulation since 1986 (*German Animal Welfare Act* as amended in 1986) and succeeds the legal principle of proportionality that asks (I) whether the proposed research is suitable in achieving a rightful goal (suitability of the research study), (II) whether the question is adequate (necessity of the research study) and (III) whether the hoped-for benefits outweigh the harms inflicted upon the animals (Luy, 2015). Question III is the actual HBA whereas questions I and II are preconditions for the HBA.

When less than optimal protocols are to be employed, for instance, when pain relieving medication is withheld, researchers must justify that the hoped-for benefits of the proposed experiments continue to outweigh the harms inflicted upon the animals. It has been a provision well before the implementation of the new European Directive, 2010/63/EU on the protection of animals used for scientific purposes in 2013, to explain in detail in the research application how the proposed research study would be conducted, including all Refinement methods that would be employed (and information that the other 2 Rs were also considered). The study findings highlight that the requirement of minimizing pain, distress, suffering and lasting harm to an absolute minimum, a provision that has existed in the law since 1972 (Section 9, para. 1, sentence 1, No.1 *German Animal Welfare Act* as amended on 24 July 1972; 1986; 1998; and current version) was frequently not fulfilled when submitting the research proposals. It also became apparent that scientific knowledge on Refinement methods that was available by 2010 was frequently not applied in practice (see e.g., knowledge on severity of procedures available through guidance documents (Paper 2) and evidence of appropriateness of certain killing methods such as CO2 and cervical dislocation (Paper 3)).

Lack of expertise in Refinement methods

Besides being familiar with and following relevant legislation, animal experimenters must be acquainted with the natural as well as pain behavior of the species they are working with, to be able to correctly assess their state of health and wellbeing. They must be proficient and properly trained to recognize and to alleviate pain (Flecknell *et al.*, 2007). Research workers are not only accountable for the research results they produce, but they are also fully responsible for the proper treatment and care of the animals in their research projects. Another key prerequisite, besides having the necessary expertise on the adequate treatment of research animals, is that principal investigators need to have skilled staff in sufficient numbers to support them in fulfilling their legal duties. Section 9, para. 2, No. 3, subset 2, which was added to the German animal welfare law in 1986, highlights that a high work load and a lack of time and money are unacceptable reasons for not providing animals with the optimal attendance and care (*German Animal Welfare Act* as amended in 1986). Directive 2010/63/EU's Article 23 on competence of personnel and Article 24 on specific requirements to personnel (which are listed in detail in Annex V of Directive 2010/63/EU) stress this basic requirement (European Parliament, 2010). Comprehensive education and ongoing training for all research workers in the 3Rs is imperative to ensure that they have the legally (and ethically) required expertise, something that was also stressed by Russell (2005) when he reviewed the progress of the 3Rs. Considering the lack of Refinement methods used that was detected in the proposals from 2010, especially the lack of pain assessment and management, training and thus expertise may still be insufficient to avoid unnecessary suffering.

A vital way to remedy these deficiencies is to intensify and improve education and training not only on all of the 3Rs of animal experimentation but on innovative, animal-free approaches in science. The European Commission's Joint Research Centre (JRC) has coordinated a study

reviewing available educational resources supporting the 3Rs principles. The goal was to find courses, teaching material and other available resources internationally and, in a next step, to produce further guidance to support educators at various levels of education (high school, university, and professional development) to include the 3Rs in curricula and educational and training programs (personal communication with Prof. Maurice Whelan, Head of Unit F.3 Chemical Safety and Alternative Methods, European Commission Joint Research Centre, May 2019) The JRC points out that knowledge sharing is an important step towards expedited development and uptake of 3Rs approaches (European Commission, 2018), with replacement methods to accelerate and aid the uptake of animal-free approaches in science being the primary focus, which also represents the first action in response to the European Citizens' Initiative "Stop Vivisection" (European Commission, 2015).

Faulty harm-benefit analyses
The underestimation of harms as well as the omissions of Refinement use contribute to incorrect prospective harm-benefit analyses by investigators and, thus, if not detected and rectified by the competent authority before licensing, conceivably lead to projects obtaining licenses in which the harms may exceed the hoped-for benefits. This is disconcerting since the HBA is the foundation to safeguard animals in science. Paper 3 (Herrmann and Flecknell, 2018b) provides examples on how we rated the severity of individual surgical procedures according to guidance documents from Switzerland (Bundesamt für Veterinärwesen, 1995) and Berlin (Arbeitskreis Berliner Tierschutzbeauftragter, 2010) that were available in 2010. When assessing severity of procedures, it is assumed that they would be conducted by a skilled surgeon who fully employs experimental Refinements. It is clear that the omission or the suboptimal use of Refinement may have an enormous impact on pain, distress and suffering and thus, overall severity. It is not only that the law dictates that suffering caused to animals in science must be reduced to an absolute minimum, but also that societal concerns about the way animals are treated demands this. Thus, if these basic requirements are not fulfilled the project should not be licensed.

Recently, in a first-of-its-kind study, the actual harms inflicted on the animals used in preclinical research for several treatment interventions were retrospectively assessed (Pound and Nicol, 2018). The authors then compared the actual harms to the actual benefits of these studies that had been systematically reviewed by Perel et. al. (2007), taking into account the animals' suffering, the probability of benefit and the importance of the research, with the last item referring to both scientific rigor and conceptual improvements. Harms were mostly rated 'severe' by an expert panel, and all studies were found to be of poor quality. Among other findings, reported use of pain medication was exceptional, and several painful procedures had been conducted with no or light anesthesia, indicating extensive shortcomings in the application of Refinement methods. The retrospective HBA, taking into account also the poor scientific rigor and the partly needless suffering of the animals, revealed that less than 7% of the animal studies were acceptable (Pound and Nicol, 2018). This retrospective HBA highlights that the oversight systems in place to safeguard animals were unsuccessful in shielding the animals from severe harm and in only authorizing high quality research that is beneficial for humans.

2.6 Significance of results within a broader context

The general lack of planned use of experimental Refinement methods that was identified in this retrospective review of original animal research proposals from 2010 indicates that there is a larger-scale problem, most probably not only in Germany. It is not only that Refinement is underreported in animal research publications (see e.g., Bara and Joffe, 2014; Carbone and Austin, 2016), but of much greater concern is that Refinement method application in practice is

likely to be poor (see Papers 1 to 3). The results of our study imply a lack of awareness and knowledge of what is legally required and how optimal Refinement can be achieved. Furthermore, the study suggests a poor understanding of the significance of Refinement to both animal wellbeing and rigor of scientific data collected.

The observation that "[s]cientists are not sufficiently aware of the concept of refinement" and "do not recognize the importance of refinement for their research" (Balls *et al.*, 1995, p. 849) was already made by the participants of the 11th European Centre for the Validation of Alternative Methods (ECVAM) Workshop in 1995. It was recommended that "[t]he concept of recognizing, minimizing and eliminating pain and distress in laboratory animals should be included in training programs for all persons involved in the care and use of laboratory animals." (Balls *et al.*, 1995, p. 849). Training programs covering the 3Rs and their application in practice were in place in 2010. However, considering the comprehensive shortcomings of the assessed proposals in that year, access to education and ongoing training may have been difficult, and/or the education and training provided may have been insufficient. Therefore, much more investment in the education and continuous training of those using animals in experiments is required. Overall, the 1995 ECVAM workshop report identified the same problems in regards to Refinement that we found in our data set from 2010.

2.7 Study limitations

Only two species were covered with the focus on various surgical interventions with subsequent recovery, and only basic and applied research proposals submitted in 2010 to most of Germany's competent authorities were included in this study. There are additional areas of concern with other animal species used to model severe disease that do not require surgeries to induce disease. Thus, the study offers a picture of a small but arguably representative segment of basic and applied animal research proposals.

Additional research would be important to track how these experiments were conducted in practice to see if all Refinements described in the proposals were actually applied, plus to see which potential additional Refinements were mandated by the competent authority in the course of the licensing process. And it would be important to check if the prospective severity classification complied with the actual severity of procedures. To conduct such retrospective assessments including reviewing and comparing prospective and actual severity, are since 2013 a new legal requirement in the EU for certain animal experiments, namely for all animal studies classified as 'severe' and for studies utilizing non-human primates (Article 39, European Commission, 2010). Hence, laboratory inspections should be included in such analyses. Furthermore, the results of these studies should be taken into account as well, for example, by reviewing the resulting publication or, if nothing is published, by investigating why the study findings were not published. This would allow lessons to be learned about best practice approaches in species- and model-specific Refinements. It would also enable lessons to be learned about the *actual* overall severity of experiments. This information is of utmost importance for future harm-benefit analyses as it facilitates a realistic consideration of the consequences for the animal subjects in the light of a more realistic estimation of potential human benefits.

2.8 Latest research, vital future research and other provisions

A survey by Franco and colleagues (2018), conducted in 2014 and 2015 among participants of laboratory animal science (LAS) courses in Denmark, Germany, Portugal, and Switzerland (n=310), found that animal researchers regarded Refinement to be more important and more

attainable than replacement. They also prioritized Refinement over Reduction efforts. This is problematic as it constitutes a reversal of the hierarchy proposed by Russell and Burch (Russell and Burch, 1959), who put Replacement first and Refinement last. This may indicate that since 2010, there has been an increase in awareness regarding the importance of Refinement use. In a follow-up survey six months after the LAS course attendants were asked to rate the impact of the themes taught in the course on their daily work. 127 participants, who answered both the first and second survey, reported in the follow-up that Refinement information taught had the most impact on their work, with over 32% saying that it had a “profound influence” and 48% stating it had “some influence” (Franco *et al.*, 2018).

A recently conducted survey (spring 2019) by Weary, Amendola and Brunt from the Animal Welfare Program, University of British Columbia, assessing opinions and attitudes of the laboratory science industry stakeholders in Canada and Europe on humane killing of laboratory rodents and on barriers to Refinements found that over half of participants from Canada and over 40% of participants from Europe answered that they still use carbon dioxide to kill rats. Most of the responders were indecisive whether this killing method is adequate or acceptable (personal communication with co-investigator Lucia Amendola, August 2019). Extensive evidence (gathered in the past 30 years and beyond) highlights that CO_2 exposure is highly aversive to the animals and must therefore be considered an inhumane killing method (for details, see Paper 3, Herrmann and Flecknell, 2018b). Many research workers seem not to be aware of existing evidence about methods causing unnecessary pain and suffering. The ongoing widespread use of CO_2 use in laboratories is a pertinent example as it highlights how animal welfare and Refinement research findings are not being transposed into practice as studies to assess the adequateness of carbon dioxide use are probably the most exhaustively conducted ones among all Refinement research today (see also Paper 3, Herrmann and Flecknell, 2018b). But instead of using other, less aversive, alternatives to carbon dioxide killing the debate is still ongoing with many continuing to argue that better alternatives were not yet available. It would be interesting to study potential conflicts of interests of those claiming that humane alternatives do not exist as evidence indicates otherwise. In a recent attempt to facilitate dialogue among a range of stakeholders, the Swiss Federal Food Safety and Veterinary Office (FSVO) organized a symposium entitled “Alternatives to Carbon Dioxide”. The compelling evidence on why CO_2 should not be used was presented as well as less aversive methods. Instead of proposing concrete action steps to initiate immediate change the FSVO concluded that more research is needed to find alternatives to carbon dioxide for killing of rats and mice (Axiak Flammer *et al.*, 2019).

Besides the necessity for research mentioned below, one future task will be the assessment of barriers to the application of Refinements, including investigations into conflicts of interest that may inhibit a change to a responsible treatment of research animals. The above-mentioned survey by Amendola *et al.* (in preparation) is part of a larger study reviewing potential and actual hurdles of implementing Refinement methods to the use of CO_2 (personal communication with Lucia Amendola, August 2019). Similar studies need to be conducted identifying the various obstacles that hamper the transposition of knowledge gained from Refinement and animal welfare research into practice.

Repetition and expansion of this study
To validate trends detected in the survey of Franco *et al.* (2018) from 2014/2015 and to assess whether the situation has changed since 2010, it will be critical to repeat our study, as well as to extend it to other species, include more areas of Refinement (e.g., housing and care), and study the situation in other countries. In addition, studies should encompass also a sample of animal protocols that are ultimately authorized after being reviewed by the competent authority and their

animal research committees to assess the work of these safeguarding bodies. Furthermore, studies may also include laboratory inspections to observe how experiments are conducted. Furthermore, the outcome of the animal studies should be analysed, especially in terms of actual severity of procedures. Such reviews should be conducted regularly, and their findings should be published, along with recommendations on how to improve practice. These could then contribute to the development of guidance documents, drafted by international experts in their respective fields, who are skilled not only in a certain area of research but also in Replacement, Reduction and Refinement of animal experiments.

Systematic documentation and assessment of actual severity
Many procedures carried out on animals, including the genetic modification of more and more animals, have introduced new challenges in appropriately assessing the severity and in classifying such procedures in the accurate categories (mild, moderate and severe) as well as to rate the cumulative effects of consecutive procedures on the individual animal. The new vistas have also provoked considerable debate. Hence, it is crucial to conduct studies alongside running research projects that fully focus on documenting and rating harms to help establish data on actual severity of the various experimental procedures and interventions. Additional animal experiments should however be avoided as this is neither necessary nor ethically justified.

Retrospective assessments of animal studies
Directive 2010/63/EU brought additional tools that ought to improve animal protection, such as mandatory retrospective assessments (RAs) of certain experiments, namely ‘severe’ projects and all projects using non-human primates (Recital 40 and Article 39, European Parliament, 2010). These retrospective severity assessments should be conducted, not only for experiments that were prospectively classified as ‘severe’ and experiments on non-human primates, but for all experiments for which the severity is unclear or/and under debate. In the course of the experiments, researchers should assess and document the ‘actual’ severity of their procedures. In the RAs, this data is then compared with their prospective severity assessments. With RAs, deficiencies and omissions in the application of all the 3Rs can be identified. Furthermore, retrospective assessments show if the prospective harm-benefit analysis was correct, by not only assessing the actual harms inflicted upon the animals but also by appraising if the study outcomes justify the inflicted harms. The conduct of mandatory retrospective assessments of projects categorized as being ‘severe’ and of projects with non-human primates was added to the list of tasks of the competent authorities (Article 39, Directive 2010/63/EU, European Commission, 2010).

Sufficiently equipped competent authorities
Considering the extent of the shortcomings found only in the area of Refinement in these research protocols, it is obvious that this places a major strain on the competent authority members. In Germany (as probably in many other countries) competent authorities are generally under-staffed and under-resourced (Personal communication with Dr. Heidemarie Ratsch, head of the competent authority in Berlin from 1996–2016, July 2016). To identify all of the shortcomings in regards to Refinement and to proper severity rating in the research applications, in the limited time the competent authority members have to assess the applications seems impossible, especially since these constitute only a part of what the competent authorities need to scrutinize when assessing animal research proposals. The brief and legally binding time limits for proposal assessments could then lead to the licensing of projects employing protocols that cause avoidable suffering and that may not be worthy a license at all.

To comply with legal requirements, competent authorities have to be adequately staffed and equipped. Besides sufficient numbers of experienced staff, there must be access available to all relevant information as well as support by external independent experts. The law dictates that, when deciding about the indispensability of animal experiments, using the current scientific knowledge as a basis, it must be assessed whether the pursued purpose cannot be reached by a means other than by the use of animals (Section 7, para. 2, sentence 3, *German Animal Welfare Act* as amended in 1986; cf. Section 7a, para. 2, sentence 1, No. 1 and 2, *German Animal Welfare Act* in its current version). However, competent authority members have no or no easy access to the full text of publications that are not freely available (Senatsverwaltung für Justiz und Verbraucherschutz Berlin, 2016). Thus, they cannot fulfill their duty in checking the literature provided in the proposals with which the researchers justify the necessity to use live animals and the indispensability of their projects per se. Competent authority members are seemingly unable to conduct a comprehensive literature search, due to not having full access to biomedical databases for reviewing the relevant full text literature, e.g. Web of Science, PubMed or National Library of Medicine (Senatsverwaltung für Justiz und Verbraucherschutz Berlin, 2016). To become informed about the current knowledge in the respective research field and about the current knowledge regarding the 3Rs, access to the relevant publications is however a central prerequisite for regulators. The competent authorities are under-resourced with respect to the country's laws and their society's concerns about animal treatment; this situation must urgently be changed.

Outsourcing the retrospective assessments
Assuming that it will take a while to increase the limited resources currently available to competent authorities - which make safeguarding animals difficult to impossible - retrospective assessments should be outsourced to independent, EU-wide working groups comprised of specialists with extensive expertise not only in the application of the 3Rs but also in the respective research field. Publication of the outcomes of such assessments would then allow for a more effective implementation of the 3Rs in the future.

Increased scrutiny by competent authorities
In Switzerland, Vogt and colleagues (2016) also assessed animal research applications submitted to the competent authorities. They screened over 1,200 proposals for several basic measures against bias (e.g., blinding, randomization and sample size calculation) and compared them with reporting rates in a sub-sample of published articles (n=50) following these study proposals. Measures to avoid bias were rarely included in both the animal research applications and in the resulting publications. The authors concluded that the authorization procedures of animal experiments and the peer review process of scientific publications are currently insufficient as they do not address the lack of study quality. This governmentally funded review led, according to a recent article (Kwon, 2019), to a more rigorous assessment of the quality of animal study applications by the Swiss competent authorities and their committees. The German government should take similar measures ensuring that the competent authorities are provided with the means of increasing the scrutiny with which they assess the animal experimentation requests - in the assessment areas of experimental Refinement and scientific quality, as well as in the other 2Rs.

Critical appraisal of the scientific value of research involving animals
The biomedical (animal) research industry is in deep crisis (Harris, 2017), due to the lack of robustness of its findings ("reproducibility crisis") and due to the lack of translational success (see Introduction and Paper 4, Herrmann, 2019).

- *The 3Vs of animal experimentation*

To assess and improve scientific validity and reproducibility of animal-based research, Würbel (2017) proposed, when conducting the harm-benefit analysis, to not only include 3Rs methods that minimize the extent and thus weight of harm inflicted upon the animals but to also incorporate the 3Vs of scientific validity (internal, external and construct validity) to maximize probable benefits of a study. The author recommends giving the 3Vs also a central role in the peer review of grant proposals and papers submitted for publication (Würbel, 2017).

- *Incomplete reporting*

Despite ample levels of endorsement of the ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) guidelines, first published in 2010 (Kilkenny *et al*., 2010) and other reporting checklists by journals and funding bodies, improvements with regard to quality and completeness of reporting in animal-based research publications have not been achieved (Hair *et al.*, 2018; Leung *et al.*, 2018). Besides other deficiencies, animal research publications fail to (fully) report whether and which refinement methods were employed. In an attempt to better this situation, Percie du Sert and colleagues (2019) recently revised the ARRIVE guidelines. On the basis of a Delphi panel that regarded 10 items as most important to evaluate the reliability of research findings, the authors decided to divide the reporting items into the "ARRIVE Essential 10", that in their opinion represented a minimum requirement, and the "Recommended Set" which contains - besides important information on "Objectives", "Ethical Statement", "Generalisability/translation" and "Interpretation/scientific implications" incl. "limitations of the animal model" all items describing Refinement method use, namely information on animal care and monitoring (description of measures taken to minimize pain, suffering and distress; any unexpected adverse events that took place, the use of humane endpoints, and the frequency of monitoring as well as information on housing and husbandry conditions, including environmental refinement). Percie du Sert *et al.* (2019) state that "[t]he intention of ARRIVE 2019 is […] to promote a harmonised approach across journals to ensure that all manuscripts contain essential information needed to appraise the research." To not include Refinement methods in the essential item list is surprising, considering the extensive research demonstrating how poor animal welfare impacts scientific data obtained from these individuals (Howard, 2013; Lloyd, Foden and Wolfensohn, 2008; Prescott and Lidster, 2017) and thus, non-application of Refinement methods makes the reliability of research findings questionable.

We recommended in Paper 1 (Herrmann and Flecknell, 2019) to legally enforce the compliance with preparation and reporting checklists, accompanied with a mandatory training on all checklist items, as this appears indispensable in fulfilling existing legal requirements. Considering the experimenters' scientific and ethical responsibility when using live animals future research on ways to improve the application of comprehensive planning and reporting guidelines that include all necessary animal welfare measures would be valuable.

- *Selective reporting of animal experiments and studies*

Moreover, research findings are often reported selectively (only the parts of the results/experimental groups that show positive results are described) (Ioannidis, 2012) and frequently, findings are not published at all which results in publication bias. Scherer and colleagues (2018) reviewed if abstracts presented at conferences resulted in a publication and found that over half of results from abstracts did not get published in full. The authors found that favorable results were more frequently published than 'negative' results (Scherer *et al.*, 2018). This selective reporting, along with other biases, leads to an overestimation of the significance of the reported study outcomes (e.g., Cohen, 2018) and thus, is misleading and contributes to

translational failures. In addition, public research funding should be used to tackle pressing health problems important to populations, and research outcomes should be relevant to patients. Surprisingly, thus far, tax-payer funded research hardly correlates with disease burden (Chalmers and Glasziou, 2009). This matter should be investigated further.

- *Increased critical appraisal of external validity*

There is an ongoing dispute as to why animal models fail to be predictive. Most argue it is due to poor scientific quality (internal validity issues) and poor reporting. However, we now mostly deal with multifactorial, complex, often chronic human illnesses of which many are not well understood and, hence, are nearly impossible to mimic in other animals (Tsukamoto, 2016). The external validity is limited due to insurmountable species differences (Pound and Bracken, 2014; Pound and Ritskes-Hoitinga, 2018), a fact that Russell and Burch already discussed six decades ago (1959, Chapter 5); many results from animals are hence not translatable to humans (Pound and Ritskes-Hoitinga, 2018). This issue is however still ignored by many in the animal research community (see e.g., Fitzpatrick *et al.*, 2018); and thus, it needs special research attention.

The scientific value of data from animals who have been altered in an attempt to mimic conditions in humans needs to be rigorously appraised through tools such as systematic reviews, meta-analyses and other retrospective assessments (as discussed in Paper 4, Herrmann, 2019). Thus far, few reviews have been undertaken looking at the value of animal experimentation. Leenaars and colleagues (2019) concluded in their recent systematic scoping review of reported concordance rates that more studies are required to evaluate the presumable evidence for predictivity of animal-based studies for the human setting and to identify which factors influence this.

2.9 Conclusions

The Three Rs Declaration of Bologna, adopted by the 3rd World Congress on Alternatives and Animal Use in the Life Sciences, Bologna, Italy, on 31 August 1999 states: “Humane science is a prerequisite for good science, and is best achieved in relation to laboratory animal procedures by the vigorous promotion and application of the Three Rs. The ‘Three Rs’ should serve as a unifying concept, a challenge, and an opportunity for reaping benefits of every kind – scientific, economic and humanitarian.” (ATLA Staff Writer, 2000).

Our Refinement study conducted with a sample of over 500 animal research proposals from Germany confirms that the insight provided by Russell and Burch 60 years ago, reaffirmed by the Three Rs Declaration of Bologna forty years later – that humane science is good science which is best achieved by effective application of the 3Rs – still seemingly has not been fully embraced by, or implemented in, many animal research laboratories. Prescott and Lidster (2017) have observed a “reluctance to question and challenge established practices and cultures” and that “staff on the ground may not have the practical tools, resources and support needed to put the current knowledge base into practice” (p. 152). These barriers to Refinement implementation have to be overcome. It is the duty of research workers to critically appraise their practices to determine current best practice approaches to avoid animal use whenever possible and to minimize welfare implications of the animals still considered indispensable while maximizing scientific results (Lloyd, Foden and Wolfensohn, 2008). “It is particularly important that this is seen as an ongoing requirement: new developments in refinement may turn what was yesterday’s best practice into today’s outdated methodology”, Lloyd and colleagues wrote (2008, p. 286). The consequences of a lack of best practice-approaches in Refinement are as obvious as they are avoidable: needless animal suffering, a breach of practical ethics, violations against animal protection laws as well as poor quality, internal validity and reproducibility of animal study

results (e.g., Prescott and Lidster, 2017). Besides overcoming the various barriers to implementing Refinement, we suggested in Paper 3 (Herrmann and Flecknell, 2018b) that international working groups with expertise in the 3Rs and their respective research field are installed, to first review the validity of the particular animal model, and, in a second step, when proven valuable, gather and provide recommendations for the Refinement of the animal model.

As Russell and Burch remarked six decades ago, “refinement is never enough, and we should always seek further for reduction and if possible replacement” (Chapter 4, 1959).To move towards Replacement and innovative, animal-free approaches appears to be the logical next step as extensive efforts have been made to optimize animal models but they have not mitigated the translational failure rate (e.g., Percie du Sert and Rice, 2014; Sutherland *et al.*, 2012). The technology revolution has already greatly changed the field of life sciences and now supplies us with novel, superior tools that make a shift away from animal experimentation increasingly feasible. The time seems right to move beyond the 60-year old 3Rs principles of animal experimentation (see Herrmann *et al.*, 2019). An ethically-changing society is increasingly demanding that concerted efforts must be made to increase the funding for the development and implementation of non-animal, human-biology based and thus human-relevant methods. Such stipulations will not only protect animals from being used in experiments but will also offer the best opportunity to advance the life sciences in their quest for cures and treatments to human illnesses.

3 Summary

Refinement on the way towards Replacement of animal experiments: A retrospective review of the use of Refinement methods in German animal research applications

Animal experimentation has been one of the most controversial areas of animal use, mainly due to the intentional harms inflicted upon animals. When applying the principles of the 3Rs (Replacement, Reduction and Refinement), first described by Russell and Burch in *The Principles of Humane Experimental Technique* six decades ago, animal use in science should decline through reducing numbers of animals used and by replacing them with non-animal methods. In addition, the use of the third R, Refinement, should lead to an improvement of the conditions that cause fear, distress, pain and suffering to animals who are still being used. It is estimated that more than 127 million vertebrates are being used in science annually. Thus, until animal subjects can be fully replaced with non-animal methods, research into the third R, Refinement, and its application in practice remains imperative. To ensure scientific rigor, avoid unnecessary animal suffering and enable an accurate harm-benefit analysis, researchers using animals must be knowledgeable about Refinement methods and apply them effectively. Indeed, researchers are legally accountable and required to minimize pain and suffering caused to the animals by their research projects.

The aims of this thesis were to evaluate if and how findings from Refinement research have been implemented in practice, to determine areas where further improvement is needed and to make recommendations for the implementation of current best-practice Refinement methods. To highlight the significance of the Replacement of animals in science for scientific and for ethical reasons, a brief overview of the use of all the 3Rs in practice has also been provided and issues of poor rigor, reproducibility and translatability of animal experimentations are discussed.

The study focused on experimental Refinement methods designed to minimize animal suffering. Research in this area has increased significantly over the past two decades. However, the extent to which Refinement methods are applied in practice is uncertain. Due to poor reporting standards in animal research publications, especially of Refinement, this study used original animal research proposals to retrospectively review planned Refinement. For the scope of the study, all applications submitted to the participating German competent authorities in 2010, that included recovery surgical procedures in mice and rats and that obtained a project license, were selected. The first paper of this publication-based thesis summarizes the findings of the assessment on the efficacy of proposed anesthesia and peri- and postoperative analgesia in animal research applications. Postoperative analgesia was not proposed for 30% of surgeries. Following 10% of procedures, animals were to be given pain analgesics if the investigators decided this was necessary; however, information on planned pain assessments were generally absent. Consequences of unalleviated pain and omission of pain assessment techniques are discussed, and recommendations to improve anesthesia and analgesia are made.

In the second paper, we categorized the severity of proposed surgical interventions by using the relevant guidance documents and compared our rating with the rating included in the proposals. Furthermore, we assessed the appropriateness of planned monitoring of the potential adverse effects on the animals (“health monitoring”). Analysis showed that severity and chronicity of pain, suffering and distress that were to be inflicted upon the animals were frequently underestimated by the researchers. Planned health monitoring generally appeared insufficient to ensure the welfare of animals in the studies.

In the third paper, the appropriateness of proposed humane endpoints and killing methods was discussed. Following 55% of surgical procedures, the investigators proposed humane endpoints. These were mostly not clearly defined, and/or were considered to be at a late stage in the animal's deterioration. In addition, the choice of killing methods could be improved in many applications, thereby reducing suffering.

The findings of this series of studies highlight the need for improvement in the implementation of Refinement in scientific use of animals, to fulfill legal requirements, to improve animal welfare and to improve the quality of research data. This study was the first to review Refinement methods in research proposals. It would be important to repeat it to assess if the situation has changed since 2010, as well as to expand it to other countries.

The final paper provides an overview of the various areas of Refinement, namely housing, husbandry, care, experimental refinements and also improvements in planning, conduct and reporting of animal studies, and their application in practice. Efforts to implement the other two Rs, Reduction and Replacement, are also briefly discussed. This paper highlights extensive shortcomings in all areas of Refinement. In addition, it supports the view that, even though we have the political goal to phase out animal experimentation, in reality, we are not close to widespread Replacement of animal models with non-animal models as not enough is done to work towards such a paradigm change.

4 Zusammenfassung

Refinement von Tierversuchen auf dem Weg zum Ersatz von Tierversuchen: Eine retrospektive Analyse der Verwendung von Refinement-Methoden in deutschen Tierversuchsanträgen

Tierversuche gehörten schon immer zu den am meisten umstrittenen Bereichen der Tiernutzung, vor allem, weil den Tieren absichtlich Schmerzen und Leiden zugefügt werden. Das 3R-Prinzip (Replacement, Reduction und Refinement), das im Deutschen auch das 3V-Prinzip (Vermeidung, Verminderung und Verfeinerung) genannt wird, wurde erstmal vor sechs Jahrzenten von Russell und Burch in ihrem Buch *The Principles of Humane Experimental Technique* beschrieben. Es sollte Tierversuche ersetzen und so den Einsatz von Tieren in der Wissenschaft reduzieren und die Bedingungen verbessern, die Schmerzen, Leiden, Distress und Angst bei den Tieren verursachen, die nach wie vor zu Versuchszwecken genutzt werden. Es wird geschätzt, dass mehr als 127 Millionen Wirbeltiere jährlich für Forschungszwecke verwendet werden. Bis Tiere vollständig durch tierfreie Methoden ersetzt werden können, werden die Forschung im Bereich Refinement und ihre Anwendung in der tierexperimentellen Praxis unerlässlich bleiben. Um die wissenschaftliche Qualität zu sichern, unnötiges Tierleid zu vermeiden und eine akkurate Schaden-Nutzen-Abwägung zu ermöglichen, müssen Tierexperimentatoren ein fundiertes Wissen über Refinement-Methoden haben, und sie müssen diese Methoden der Leidensminimierung erfolgreich anwenden können. In der Tat sind Wissenschaftler rechenschaftspflichtig und gesetzlich verpflichtet, die den Tieren durch ihre Forschungsprojekte zugefügten Schmerzen und Leiden auf ein unerlässliches Maß reduzieren.

Die Ziele dieser Arbeit waren zu evaluieren, ob und wie die Erkenntnisse aus der Refinement-Forschung in der Praxis implementiert worden sind, zu untersuchen, wo Verbesserungen notwendig sind und Empfehlungen für die weitere Optimierung der Anwendung von Refinement-Forschungsergebnissen in der Praxis zu geben. Um zu verdeutlichen, dass das endgültige Ziel der Ersatz von Tieren in der Wissenschaft sein sollte – aus ethischen wie aus wissenschaftlichen Gründen – wird eine Übersicht über die Anwendung aller drei R in der Praxis vorgestellt, und es werden Mängel bezüglich der Qualität, Reproduzierbarkeit und Übertragbarkeit von Tierversuchen diskutiert.

Die Arbeit konzentrierte sich auf experimentelle Refinement-Methoden, die entwickelt wurden, um versuchsbedingtes Tierleid zu minimieren. Die Forschung in diesem Bereich hat in den letzten zwei Jahrzehnten stark zugenommen. Aber es ist unklar, in welchem Umfang Refinement-Methoden in der täglichen Tierversuchspraxis Anwendung finden. Aufgrund mangelhafter Standards bei der Berichterstattung von tierexperimentellen Studien, insbesondere bezüglich ihrer Verwendung von Refinement-Maßnahmen, wurden für diese Studie Tierversuchsanträge herangezogen, anhand welcher die geplanten Refinement-Maßnahmen untersucht wurden. Es wurden alle Tierversuchsanträge mit operativen Eingriffen an Mäusen und Ratten, die im Jahr 2010 an die an der Studie teilnehmenden zuständigen Behörden für Tierversuche gesendet wurden und die schlussendlich eine Genehmigung erhielten, ausgewählt.

Der erste Artikel dieser publikationsbasierten Arbeit fasst die Ergebnisse der Untersuchung zur Effektivität der geplanten Anästhesie- und peri-und postoperativen Analgesieregime zusammen. Nach 30% der operativen Eingriffe war keine postoperative Analgesie geplant. Bei 10% der Operationen gaben die Tierexperimentatoren an, eine postoperative Schmerzbehandlung zu machen, falls sie dies für notwendig hielten. Jedoch fehlten in den Anträgen Hinweise auf eine geplante Beobachtung und Bewertung von Schmerzen und Leiden der Versuchstiere. Die

Konsequenzen von fehlendem bzw. mangelhaftem Schmerzmanagement werden diskutiert, und es werden Empfehlungen zur Verbesserung der Anästhesie- und Analgesieprotokolle gemacht.

In der zweiten Publikation haben wir die geplanten operativen Eingriffe in Schweregrade eingeteilt und unsere Bewertung mit der der Antragsteller verglichen. Außerdem überprüften wir, wie geeignet die vorgesehene Überwachung des Gesundheitszustandes der Tiere ist. Unsere Analyse ergab, dass Schweregrad und Chronizität von Schmerzen, Leiden und Dysstress, denen die Tiere ausgesetzt werden würden, von den Forschern regelmäßig unterschätzt wurden. Außerdem erschien die geplante Überwachung des Gesundheitszustandes der Tiere unzureichend, um das Wohlergehen der Tiere in diesen Versuchen sicherzustellen.

Der dritte Artikel diskutiert die Angemessenheit der von den Antragstellern vorgeschlagenen humanen Endpunkte (Abbruchkriterien) und Tötungsmethoden. Wenn die Forscher Abbruchkriterien im Antrag genannt hatten, was nach 55% der chirurgischen Eingriffe der Fall war, waren diese meist nicht eindeutig definiert und/oder wurden im Hinblick auf den bereits stark verschlechterten Krankheitszustand der Tiere als zu spät erachtet. Darüber hinaus könnte die Wahl der Tötungsmethode in vielen Anträgen verbessert werden, was zur Leidensminimierung beitragen würde.

Die Ergebnisse dieser Studienreihe machen die Notwendigkeit für Verbesserungen bei der Anwendung von Refinement-Methoden in der alltäglichen tierexperimentellen Praxis deutlich – um gesetzliche Vorgaben zu erfüllen, um den Tierschutz zu verbessern und um die Qualität der generierten Daten zu optimieren. Diese Studie war die erste, die Refinement-Methoden in Tierversuchsanträgen analysiert hat. Es wäre wichtig, diese Studie zu wiederholen, um zu eruieren, ob sich die Situation seit 2010 verändert hat; außerdem wäre es wichtig, derartige Untersuchungen auf andere Länder auszuweiten.

Der letzte Artikel gibt einen Überblick über derzeitige wissenschaftliche Erkenntnisse in den verschiedenen Bereichen des Refinements, und zwar über die optimierte Unterbringung und Pflege, über experimentelles Refinement und über Verbesserungen in der Planung, Durchführung und Berichterstattung von Tierversuchen. Es wird außerdem dargelegt, wie diese Erkenntnisse derzeit in der Praxis umgesetzt werden. Bemühungen, die anderen 2R, Reduzierung und Ersatz von Tierversuchen, voranzutreiben werden auch kurz diskutiert. Es werden umfangreiche Mängel in allen Bereichen des Refinements von Tierversuchen und ihrer Anwendung in der Praxis aufgezeigt. Außerdem wird deutlich, dass wir trotz des politischen Ziels Tierversuche einzuschränken, in Wahrheit weit entfernt sind von einem weitgehenden Ersatz von Tiermodellen durch tierfreie Modelle und Methoden, da nicht genug getan wird um einen Paradigmenwechsel voranzutreiben.

5 Reference list of introduction and discussion

ATLA Staff Writer (2000). The Three Rs Declaration of Bologna. Reduction, Refinement and Replacement Alternatives and Laboratory Animal Procedures. Adopted by the 3rd World Congress on Alternatives and Animal Use in the Life Sciences, Bologna, Italy, on 31 August 1999. *ATLA-Alternatives to Laboratory Animals*, 28, pp. 1-5.

Anton, F., Euchner, I. and Handwerker, H.O. (1992). Psychophysical examination of pain induced by defined CO2 pulses applied to the nasal mucosa. *Pain*, 49(1), pp. 53-60.

Arbeitskreis Berliner Tierschutzbeauftragter (2010). *Orientierungshilfe des Arbeitskreises Berliner Tierschutzbeauftragter zur Einstufung in Belastungsgrade*. Available at: http://www.gv-solas.de/fileadmin/user_upload/pdf_publikation/Orientierungshilfe_21-09-10.pdf [Accessed 12 August 2018].

Archibald, K., Tsaioun, K., Kenna, J.G. and Pound, P. (2018). Better science for safer medicines: the human imperative. *Journal of the Royal Society of Medicine*, 111(12), pp. 433-438.

Axiak Flammer, S., Eskes, C., Kohler, I., Ochieng Pernet, A., Jakob, P.; Marahrens, M., Gent, T.C., Golledge, H. and Weary, D. (2019). Alternatives to Carbon Dioxide—Taking Responsibility for Humanely Ending the Life of Animals. *Animals*, 9, p. 482. Available at: https://www.mdpi.com/2076-2615/9/8/482 [Accessed 25 July 2019].

Bailey, J. (2018). Does the stress of laboratory life and experimentation on animals adversely affect research data? A critical review. *Alternatives to laboratory animals: ATLA-Alternatives to Laboratory Animals*, 46(5), pp. 291-305.

Bailey, J. (2019). Genetic modification of animals: scientific and ethical issues. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Brill Human Animal Studies, Vol. 22, Leiden: Brill, pp. 443-479.Availabel at: https://brill.com/view/title/35072?lang=en [Accessed 1 April 2019].

Baker, M. (2016). 1,500 Scientists Lift the Lid on Reproducibility. *Nature News*, 533(7604), p. 452.

Balls, M. (1999). The biomedical sciences and the need for less inhumane animal procedures. In: C.F.M. Hendriksen and D.B. Morton, eds., *Humane Endpoints in Animal Experiments for Biomedical Research*, pp. 1–4. London, UK: Royal Society of Medicine Press.

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

Banzett, R.B. and Moosavi, S.H. (2001). Dyspnea and pain: similarities and contrasts between two very unpleasant sensations. *APS Bulletin*, 11(1), pp. 1-8.

Banzett, R.B., Pedersen, S.H., Schwartzstein, R.M. and Lansing, R.W. (2008). The affective dimension of laboratory dyspnea: air hunger is more unpleasant than work/effort. *American Journal of Respiratory and Critical Care Medicine*, 177(12), pp. 1384-1390.

Bara, M. and Joffe, A.R. (2014). The ethical dimension in published animal research in critical care: the public face of science. *Critical Care*, 18(1), p. R15. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4056799/ [Accessed 12 July 2018].

Basketter, D.A., Clewell, H., Kimber, I., Rossi, A., Blaauboer, B.J., Burrier, R., Daneshian, M., Eskes, C., Goldberg, A. and Hasiwa, N. (2012). A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing-t4 report. *ALTEX - Alternatives to Animal Experimentation*, 29(1), pp. 3-91.

Bayne, K., Ramachandra, G.S., Rivera, E.A. and Wang, J. (2015). The Evolution of Animal Welfare and the 3Rs in Brazil, China, and India. *Journal of the American Association for Laboratory Animal Science*, 54(2), pp. 181-191.

Begley, C. G. and Ellis, L. M. (2012). Drug Development: Raise Standards for Preclinical Cancer Research. *Nature*, 483(7391), pp. 531–533.

Begley, C.G. and Ioannidis, J.P. (2015). Reproducibility in science: improving the standard for basic and preclinical research. *Circulation Research*, 116(1), pp.116-126. Available at: https://www.ahajournals.org/doi/pdf/10.1161/CIRCRESAHA.114.303819 [Accessed 12 May 2018].

Bertrand, H.G.M.J., Sandersen, C. and Flecknell, P.A. (2018). Reported analgesic and anaesthetic administration to non-human primates undergoing experimental surgical procedure: 2010-2015. *Journal of Medical Primatology*, *47*(4), pp. 217-225.

BioMed21 Collaboration (n.d.). About us. Available at: http://biomed21.org/about-us/ [Accessed 23 May 2018].

Brofman, J.D., Leff, A.R., Munoz, N.M., Kirchhoff, C.A.R.O.L. and White, S.R. (1990). Sympathetic secretory response to hypercapnic acidosis in swine. *Journal of Applied Physiology*, 69(2), pp. 710-717.

Bundesamt für Veterinärwesen (1995). Einteilung von Tierversuchen nach Schweregraden vor Versuchsbeginn (Belastungskategorien). *Information Tierschutz*, 1.04. 20pp. Available at: https://www.medizin.uni-tuebingen.de/tierschutz/116104_Belastungskatalog_schweiz_7-11_PAM.pdf [Accessed 12 May 2018].

Bundesministerium für Ernährung und Landwirtschaft (2014). Anzahl der für Versuche und andere wissenschaftliche Zwecke verwendeten Wirbeltiere. Stand 01.12.2014. http://www.bmel.de/SharedDocs/Downloads/Tier/Tierschutz/2013-TierversuchszahlenGesamt.pdf?__blob=publicationFile [Accessed 12 August 2018].

Bundesministerium für Ernährung und Landwirtschaft (2017). Versuchstierdaten 2016. Available at: https://www.bmel.de/SharedDocs/Downloads/Tier/Tierschutz/Versuchstierdaten2016.pdf?__blob=publicationFile [Accessed 12 August 2018].

Carbone, L. and Austin, J. (2016). Pain and laboratory animals: publication practices for better data reproducibility and better animal welfare. *PloS ONE*, 11(5), p. e0155001. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155001 [Accessed 12 August 2018].

Chalmers, I. and Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, 374(9683), pp. 86-89.

Clemence, M. and Leaman, J. (2016). Public Attitudes to Animal Research in 2016. A report by Ipsos MORI for the Department for Business, Energy & Industrial Strategy, *Ipsos MORI Social Research Institute*. Available at: https://www.ipsos.com/sites/default/files/publication/1970-01/sri-public-attitudes-to-animal-research-2016.pdf [Accessed 12 August 2018].

Cohen, D. (2018). Oxford TB Vaccine Study Calls into Question Selective Use of Animal Data. *British Medical Journal*, 360, p. j5845.

Council of Europe (1986). Convention for the Protection of Vertebrate Animals Used for Experimental and other Scientific Purposes. *European Treaty Series No. 123*. Available at: https://rm.coe.int/168007a67b [Accessed 12 August 2018].

Daneshian, M., Busquet, F., Hartung, T. and Leist, M. (2015). Animal Use for Science in Europe. *ALTEX - Alternatives to Animal Experimentation*, 32, pp. 261-274. Available at: https://www.altex.org/index.php/altex/article/view/208/211 [Accessed 12 August 2018].

Danneman, P.J., Stein, S. and Walshaw, S.O. (1997). Humane and practical implications of using carbon dioxide mixed with oxygen for anesthesia or euthanasia of rats. *Laboratory Animal Science*, 47(4), pp. 376-385.

EFSA AHAW Panel (2005). Opinion of the Scientific Panel on Animal Health and Welfare on a request of the Commission related to “Aspects of the biology and welfare of animals used for experimental and other scientific purposes”. EFSA-Q-2004-105. *The EFSA Journal,* 292, pp. 1-46.

European Citizen’s Initiative (2016). *Stop Vivisection*. Available at: http://www.stopvivisection.eu/de [Accessed 12 August 2018].

European Commission (2011). Seventh report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union. 52013DC0859. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52013DC0859 [Accessed 10 July 2018].

European Commission (2013). Report from the Commission to the Council and the European Parliament. Seventh Report on the Statistics on the Number of Animals used for Experimental and other Scientific Purposes in the Member States of the European Union. *COM* (2013) 859 final. Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:e99d2a56-32fc-4f60-ad69-61ead7e377e8.0001.03/DOC_1&format=PDF [Accessed 10 July 2018].

European Commission (2015). Communication from the Commission on the European Citizens' Initiative "Stop Vivisection". *C(2015) 3773 final*, pp. 1-14. Available at: https://ec.europa.eu/environment/chemicals/lab_animals/pdf/vivisection/en.pdf [Accessed 27 February 2018].

European Commission (2018). Mapping education and training on the 3Rs. Available at: https://ec.europa.eu/jrc/en/science-update/education-and-training-3rs [Accessed 27 February 2018].

European Commission (2019). The European Citizens' Initiative. Available at: https://ec.europa.eu/citizens-initiative/public/basic-facts [Accessed 27 February 2019].

European Parliament (2010). 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. *Official Journal of the European Communities*, L276, pp. 33-79. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32010L0063 [Accessed 23 June 2018].

Finlay, C.G. and Forsyth, J.P. (2009). Context and renewal of conditioned fear: an experimental evaluation using 20% carbon dioxide-enriched air as an unconditioned stimulus. *Journal of Anxiety Disorders*, 23(6), pp. 737-745.

Fitzpatrick, B.G., Koustova, E. and Wang, Y. (2018). Getting personal with the" reproducibility crisis": interviews in the animal research community. *Lab Animal*, 47(7), pp.175-177.

Flecknell, P., Gledhill, J. and Richardson, C. (2007). Assessing animal health and welfare and recognising pain and distress. *ALTEX – Alternatives to Animal Experimentation*, 24(8). Available at: https://pdfs.semanticscholar.org/6c00/e65eeaa8b0332cd1d739952cb7251ec2faf0.pdf [Accessed 23 June 2018].

Franco, N.H., Sandøe, P. and Olsson, I.A.S. (2018). Researchers' attitudes to the 3Rs—An upturned hierarchy?. *PloS ONE*, 13(8), p. e0200895. Available at: https://doi.org/10.1371/journal.pone.0200895 [Accessed 17 August 2018].

Freed, D.L.J. (1983). CO2 euthanasia. *Nature*, 304, p. 482. Available at: https://www.nature.com/articles/304482c0.pdf [Accessed 23 June 2018].

Freedman, L.P., Cockburn, I.M., Simcoe, T.S. (2015). The Economics of Reproducibility in Preclinical Research. *PLoS Biology*, 13(6), p. e1002165. Available at: https://doi.org/10.1371/journal.pbio.1002165 [Accessed 23 June 2018].

Freedman, L.P., Venugopalan, G. and Wisman, R. (2017). Reproducibility2020: progress and priorities. *F1000Research*, 6, p. 604. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5461896/pdf/f1000research-6-12234.pdf [Accessed 23 June 2018].

General Administrative Provision for the Execution of the Animal Welfare Act of 9 February 2000 (Allgemeine Verwaltungsvorschrift zur Durchführung des Tierschutzgesetzes vom 9 Februar 2000. Available at: http://www.verwaltungsvorschriften-im-internet.de/bsvwvbund_09022000_32135220 006.htm [Accessed 17 November 2018].

German Animal Welfare Act as amended on 24 July 1972 (Deutsches Tierschutzgesetz in der Fassung vom 24. Juli 1972). Available at: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl172s1277.pdf#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl172s1277.pdf%27%5D__1536251890589 [Accessed 12 August 2018].

German Animal Welfare Act as amended in 1986 (Deutsches Tierschutzgesetz in der Fassung von 1986). Available at: http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl186s1319.pdf [Accessed 12 August 2018].

German Animal Welfare Act in its current version (Tierschutzgesetz in der Fassung der Bekanntmachung vom 18. Mai 2006 (BGBl. I S. 1206, 1313), das zuletzt durch Artikel 141 des Gesetzes vom 29. März 2017 (BGBl. I S. 626) geändert worden ist). Available at: https://www.gesetze-im-internet.de/tierschg/TierSchG.pdf [Accessed 12 August 2018].

Greek, R. and Kramer, L. (2019). How to evaluate the science of non-human animal use in biomedical research and testing: A proposed format for debate. In: K. Herrmann and K. Jayne, eds. *Animal Experimentation: Working Towards a Paradigm Change*, Brill Human Animal Studies, Vol. 22, Leiden: Brill, pp. 65-87. Available at: https://brill.com/view/title/35072?lang=en [Accessed 1 April 2019].

Hair, K., Macleod, M. R., Sena, E. S. and IICARus Collaboration (2018). A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *bioRxiv*, p. 370874.

Harris, R. (2017). *Rigor mortis. How sloppy science creates worthless cures, crushes hope, and wastes billions*. New York: Basic.

Hawkins, P., Playle, L., Golledge, H., Leach, M., Banzett, R., Coenen, A., Cooper, J., Danneman, P., Flecknell, P., Kirkden, R. and Niel, L. (2006). *Report from the Newcastle consensus meeting on carbon dioxide euthanasia of laboratory animals*. University of Newcastle upon Tyne, UK, pp. 1-17. Available at: https://www.nc3rs.org.uk/sites/default/files/documents/Events/First%20Newcastle%20consensus%20meeting%20report.pdf [Accessed 12 June 2018].

Herrmann, K. (2019). Refinement on the way towards replacement: Are we doing what we can?. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Vol. 22, Leiden: Brill, pp. 3-64. Available at: https://brill.com/view/book/edcoll/9789004391192/BP000002.xml [Accessed 1 April 2019].

Herrmann, K. and Flecknell, P.A. (2019). Retrospective review of anesthetic and analgesic regimens used in animal research proposals. *ALTEX - Alternatives to Animal Experimentation*, 36(1), pp. 65-80. Available at: https://doi.org/10.14573/altex.1804011 [Accessed 12 January 2019].

Herrmann, K. and Flecknell, P.A. (2018a). Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals – a retrospective review. *ATLA - Alternatives to Laboratory Animals*, 46(5), pp. 273-289. Available at: https://www.researchgate.net/publication/329324086_Severity_classification_of_surgical_procedures_and_application_of_health_monitoring_strategies_in_animal_research_proposals_A_retrospective_review [Accessed 12 December 2018].

Herrmann, K. and Flecknell, P.A. (2018b). Application of humane endpoints and humane killing methods in animal research applications – a retrospective review. *ATLA - Alternatives to Laboratory Animals*, 46(6), pp. 317-333. Available at: https://www.researchgate.net/publication/330502849_The_application_of_humane_endpoints_a

nd_humane_killing_methods_in_animal_research_proposals_A_retrospective_review [Accessed 12 January 2019].

Herrmann, K., Pistollato, F. and Stephens, M. (2019) Beyond the 3Rs: Expanding the use of human-relevant replacement methods in biomedical research, *ALTEX - Alternatives to Animal Experimentation*, 36(3), pp. 343-352. Available at: https://www.altex.org/index.php/altex/article/view/1301/1324 [Accessed 19 July 2019].

Hooijmans, C.R., Leenaars, M. and Ritskes-Hoitinga, M. (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), pp. 167-182. Available at: http://repository.ubn.ru.nl/bitstream/handle/2066/89153/89153.pdf [Accessed 14 August 2018].

Howard, B. (2013). Russell and Burch's definition of refinement. *ATLA - Alternatives to Laboratory Animals*, 41, pp. P30-P32.

Ioannidis, J.P.A. (2012). Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science*, 7(6), pp. 645-654. Available at: http://journals.sagepub.com/doi/full/10.1177/1745691612464056 [Accessed 15 July 2017].

Johnson, P.L., Fitz, S.D., Hollis, J.H., Moratalla, R., Lightman, S.L., Shekhar, A. and Lowry, C.A. (2011). Induction of c-Fos in 'panic/defence'-related brain circuits following brief hypercarbic gas exposure. *Journal of Psychopharmacology*, 25(1), pp. 26-36.

Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M. and Altman, D.G. (2010). Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLoS Biology*, 8(6), p. e1000412. Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000412 [Accessed 14 May 2017].

Knight, A. (2008). 127 million non-human vertebrates used worldwide for scientific purposes in 2005. *ATLA - Alternatives to Laboratory Animals*, 36(5), pp. 494-496.

Kramer, L.A. and Greek, R. (2018). Human stakeholders and the use of animals in drug development. *Business and Society Review,* 123, pp. 3-58.

Kwon, D. (2019). Swiss Researchers Struggle to Get Animal Experiments Approved. *The Scientist*. Available at: https://www.the-scientist.com/news-opinion/swiss-researchers-struggle-to-get-animal-experiments-approved--65293 [Accessed 10 January 2019].

Langley, G.R., Adcock, I.M., Busquet, F., Crofton, K.M., Csernok, E., Giese, C., Heinonen, T., Herrmann, K., Hofmann-Apitius, M., Landesmann, B. and Marshall, L.J. (2017). Towards a 21st-century roadmap for biomedical research and drug discovery: consensus report and recommendations. *Drug Discovery Today*, 22(2), pp. 327-339.

Leenaars, C.H., Kouwenaar, C., Stafleu, F.R., Bleich, A., Ritskes-Hoitinga, M., De Vries, R.B. and Meijboom, F.L. (2019). Animal to human translation: a systematic scoping review of reported concordance rates. *Journal of Translational Medicine*, 17(1), p. 223. Available at: https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-019-1976-2 [Accessed 16 July 2019].

Leung, V., Rousseau-Blass, F., Beauchamp, G. and Pang, D.S.J. (2018). ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of *in vivo* Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS ONE*, 13(5): e0197882. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0197882 [Accessed 27 May 2018].

Lloyd, M.H., Foden, B.W. and Wolfensohn, S.E. (2008). Refinement: promoting the three Rs in practice. *Laboratory Animals*, 42(3), pp. 284-293.

Luy, J. (2015). The principle of proportionality: the concept and its application to animal protection issues such as the 'three Rs' and the 'harm–benefit analysis'. *ALTEX – Alternatives to Animal Experimentation Proceedings*, 4, pp. 16-23. Available at: http://www.altex.ch/resources/altex_2015_Proc1_016_023_Luy1.pdf [Accessed 16 February 2017].

Marquardt, N., Feja, M., Hünigen, H., Plendl, J., Menken, L., Fink, H. and Bert, B. (2018). Euthanasia of laboratory mice: Are isoflurane and sevoflurane real alternatives to carbon dioxide?. *PloS ONE* 13(9), p. e0203793. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0203793 [Accessed 15 September 2018].

Metschke, R. und Wellbrock, R. (2002). *Datenschutz in Wissenschaft und Forschung*. (Data Protection in Science and Research). Berliner Beauftragter für Datenschutz und Informationsfreiheit, 3rd revised edition.

Pawlowski, J., Feinstein, D., Crandall, M.L. and Gala, S. (2019). Modernizing biomedical training: replacing live animal laboratories with human simulation. In: K. Herrmann and K. Jayne, eds., *Animal Experimentation: Working Towards a Paradigm Change*, Brill Human Animal Studies, Vol. 22, Leiden: Brill, pp. 551-566. Available at: https://brill.com/view/title/35072?lang=en [Accessed 1 April 2019].

Peppel, P. and Anton, F. (1993). Responses of rat medullary dorsal horn neurons following intranasal noxious chemical stimulation: effects of stimulus intensity, duration, and interstimulus interval. *Journal of Neurophysiology*, 70(6), pp. 2260-2275.

Percie du Sert, N. and Rice, A.S.C. (2014). Improving the translation of analgesic drugs to the clinic: animal models of neuropathic pain. *British Journal of Pharmacology*, 171(12), pp. 2951-2963. Available at: https://bpspubs.onlinelibrary.wiley.com/doi/full/10.1111/bph.12645%4010.1111/%28ISSN%291476-5381.AnimalModelsinPharmacologyResearchupdatedJanuary2015 [Accessed 2 April 2018].

Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M.T., Baker, M., Browne, W.J., Clark, A., Cuthill, I.C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S.T., Howells, D.W., Karp, N.A., Lidster, K., MacCallum, C.J., MacLeod, M., Peterson, O., Rawle, F., Reynolds, P., Rooney, K., Sena, E.S., Silberberg, S.D., Steckler, T. and Würbel, H. (2019). The ARRIVE guidelines 2019: updated guidelines for reporting animal research. *BioRxiv*, p.703181. Available at: https://www.biorxiv.org/content/biorxiv/early/2019/07/15/703181.full.pdf [Accessed 16 July 2019].

Perel, P., Roberts, I., Sena, E., Wheble, P., Briscoe, C., Sandercock, P., Macleod, M., Mignini, L. E., Jayaram, P. and Khan, K. S. (2007). Comparison of Treatment Effects Between Animal

Experiments and Clinical Trials: Systematic Review. *British Medical Journal*, 334, pp. 197-206. Available at: https://www.bmj.com/content/334/7586/197 [Accessed 16 January 2018].

Pharmaceutical Research and Manufacturers of America (PhRM) (2015). *Biopharmaceutical Research Industry Profile.* Washington, DC: PhRMA. http://phrma-docs.phrma.org/sites/default/files/pdf/2015_phrma_profile.pdf [Accessed 10 June 2018].

Pharmaceutical Research and Manufacturers of America (PhRM) (2016). *Biopharmaceutical Research Industry Profile.* Washington, DC: PhRMA. http://phrma-docs.phrma.org/sites/default/files/pdf/biopharmaceutical-industry-profile.pdf [Accessed 10 June 2018].

Pippin, J.J. (2012). Animal research in medical sciences: Seeking a convergence of science, medicine, and animal law. *South Texas Law Review,* 54, pp. 469-511. Available at: http://faculty.smu.edu/jkazez/ar13/pippin. Pdf [Accessed 10 May 2018].

Prescott, M. J. and Lidster, K. (2017). Improving quality of science through better animal welfare: the NCR3Rs strategy. *Lab Animal*, 46(4), pp. 152-156.

Pound, P. and Bracken, M.B. (2014). Is Animal Research Sufficiently Evidence Based To Be a Cornerstone of Biomedical Research?. *British Medical Journal*, 348, p. g3387.

Pound, P., Ebrahim, S., Sandercock, P., Bracken, M.B. and Roberts, I. (2004). Where Is the Evidence That Animal Research Benefits Humans?. *British Medical Journal*, 328(7438), pp. 514-517.

Pound, P. and Nicol, C.J. (2018). Retrospective harm benefit analysis of pre-clinical animal research for six treatment interventions. *PLoS ONE*, 13(3), p. e0193758. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193758 [Accessed 11 June 2018].

Pound, P. and Ritskes-Hoitinga, M. (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*(1), p. 304. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6223056/ [Accessed 9 November 2018].

Prinz, F., Schlange, T. and Asadullah, K. (2011). Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?. *Nature Reviews Drug Discovery*, 10(9), pp. 712-712.

Richardson, C.A. and Flecknell, P.A. (2005). Anaesthesia and Postoperative Analgesia Following Experimental Surgery in Laboratory Rodents: Are We Making Progress?. *ALTA - Alternatives to Laboratory Animals*, 33, pp. 119-127.

Russell, W.M.S. (2005). The Three Rs: Past, present and future. *Animal Welfare*,14, pp. 279–286.

Russell, W.M.S. and Burch, R.L. (1959). *The Principles of Humane Experimental Technique.* Potters Bar, Hertfordshire, England: Universities Federation for Animal Welfare. Available at: http://altweb.jhsph.edu/pubs/books/humane_exp/het-toc [Accessed 21 January 2018].

Scherer, R.W., Meerpohl, J.J, Pfeifer, N., Schmucker, C., Schwarzer, G. and von Elm, E. (2018). Full publication of results initially presented in abstracts.
Cochrane Database of Systematic Reviews, 11. Art. No.: MR000005. Available at: https://www.researchgate.net/profile/Christine_Schmucker/publication/329226050_Full_publication_of_results_initially_presented_in_abstracts/links/5c051329458515ae54422d61/Full-publication-of-results-initially-presented-in-abstracts.pdf [Accessed 21 November 2018].

Senatsverwaltung für Justiz und Verbraucherschutz Berlin (2016). *Schriftliche Anfrage der Abgeordenten Claudia Hämmerling vom 7. März 2016 und Antwort. Tierversuchsanträge 2014 und 2015.* Abgeordnetenhaus Berlin, Drucksache 17/18169. Available at: http://pardok.parlament-berlin.de/starweb/adis/citat/VT/17/SchrAnfr/s17-18169.pdf [Accessed 13 April 2018].

Sutherland, B.A., Minnerup, J., Balami, J.S., Arba, F., Buchan, A.M. and Kleinschnitz, C., 2012. Neuroprotection for ischaemic stroke: translation from the bench to the bedside. *International Journal of Stroke*, 7(5), pp. 407-418. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1747-4949.2012.00770.x [Accessed 12 May 2018].

Taylor, K. (2010). Reporting the implementation of the Three Rs in European primate and mouse research papers: are we making progress?. *ATLA-Alternatives to Laboratory Animals*, 38(6), pp. 1-23.

Taylor, K. and Rego, L. (2016). EU statistics on Animal Experiments for 2014. *ALTEX - Alternatives to Animal Experimentation Proceedings*, 33(4), pp. 465-468.

Taylor, K. and Rego Alvarez, L. (2019). A summary of EU national statistical reports of animal experiments in 2014-2016. ALTEX-Alternatives to Animal Experimentation 36, pp. 314-319. Available at: https://www.altex.org/index.php/altex/article/view/1233 [Accessed 11 April 2019].

Tsukamoto, T. (2016). Animal Disease Models for Drug Screening: The Elephant in the Room? *Drug Discovery Today*, 21, pp. 529-530.

Uhlig, C., Krause, H., Koch, T., de Abreu, M.G. and Spieth, P.M (2015). Anesthesia and monitoring in small laboratory mammals used in anesthesiology, respiratory and critical care research: a systematic review on the current reporting in top-10 impact factor ranked journals. *PloS ONE*, 10(8), p. e0134205. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134205 [Accessed 12 May 2018].

Vogt, L., Reichlin, T.S., Nathues, C. and Würbel, H. (2016). Authorization of animal experiments is based on confidence rather than evidence of scientific rigor. *PLoS Biology*, *14*(12), p. e2000598. Available at: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.2000598 [Accessed 12 May 2018].

von Dehn, G. and Nobiling, R. (2014). Kommissionsarbeit nach Paragraph 15 Tierschutzgesetz. Eine ethische und wissenschaftliche Herausforderung. *Deutsches Tierärzteblatt*, 3/2014, pp. 324-329.

World Organisation for Animal Health (OIE) (2018). *Terrestrial Animal Health Code*, Ch. 7.8. Paris: World Organisation for Animal Health, first adopted in 2010; most recent update adopted in 2013. [online] Available at: http://www.oie.int/index.php?id=169&L=0&htmfile=chapitre_aw_research_education.htm [Accessed 12 May 2018].

Würbel, H. (2007). Publications should include an animal-welfare section. *Nature*, 446(7133), p. 257. Available at: https://www.nature.com/articles/446257a [Accessed 10 June 2018].

Würbel, H. (2017). More Than 3Rs: The Importance of Scientific Validity for Harm-benefit Analysis of Animal Research. *Lab Animal*, 46(4), pp. 164-166. Available at: https://www.nature.com/articles/laban.1220.pdf [Accessed 12 June 2018].

Ziemann, A.E., Allen, J.E., Dahdaleh, N.S., Drebot, I.I., Coryell, M.W., Wunsch, A.M., Lynch, C.M., Faraci, F.M., Howard, M.A., Welsh, M.J. & Wemmie, J.A. (2009). The amygdala is a chemosensor that detects carbon dioxide and acidosis to elicit fear behavior. *Cell*, 139(5), pp. 1012-1021.

6 List of publications

6.1 Original scientific papers

Herrmann, K. and Flecknell, P.A. (2019). Retrospective review of anesthetic and analgesic regimens used in animal research proposals. Alternatives to Animal Experimentation, 36(1), pp. 65-80. Available at: https://doi.org/10.14573/altex.1804011

Herrmann, K. and Flecknell, P.A. (2018a). Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals – a retrospective review. Alternatives to Laboratory Animals, 46(5), pp. 273-289. Available at: https://www.researchgate.net/publication/329324086_Severity_classification_of_surgical_procedures_and_application_of_health_monitoring_strategies_in_animal_research_proposals_A_retrospective_review

Herrmann, K. and Flecknell, P.A. (2018b). Application of humane endpoints and humane killing methods in animal research applications – a retrospective review. Alternatives to Laboratory Animals, 46(6), pp. 317-333. Available at: https://www.researchgate.net/publication/330502849_The_application_of_humane_endpoints_and_humane_killing_methods_in_animal_research_proposals_A_retrospective_review

Herrmann, K. (2019). Refinement on the way towards replacement: Are we doing what we can?. In: K. Herrmann and K. Jayne, eds. Animal Experimentation: Working Towards a Paradigm Change, Vol. 22, Leiden: Brill, pp. 3-64. Available at: https://brill.com/view/book/edcoll/9789004391192/BP000002.xml

6.1.1 Contribution to original scientific papers

Retrospective review of anesthetic and analgesic regimens used in animal research proposals

Kathrin Herrmann
Roles: Conceptualization, Resources, Data curation, Design, Methodology, Analysis, Investigation, Resources, Writing – original draft, Writing – review and editing

Prof. Paul Flecknell
Roles: Supervision; Review and editing

Severity classification of surgical procedures and application of health monitoring in animal research proposals – a retrospective review

Kathrin Herrmann
Roles: Conceptualization, Resources, Data curation, Design, Methodology, Analysis, Investigation, Writing – original draft, Writing – review and editing

Prof. Paul Flecknell
Roles: Supervision; Review and editing

Application of humane endpoints and humane killing methods in animal research applications – a retrospective review

Kathrin Herrmann
Roles: Conceptualization, Resources, Data curation, Design, Methodology, Analysis, Investigation, Writing – original draft, Writing – review and editing

Prof. Paul Flecknell
Roles: Supervision; Review and editing

Refinement on the way towards replacement: Are we doing what we can?

Single authorship: Kathrin Herrmann
Roles: Conceptualization, Resources, Design, Methodology, Analysis, Writing – original draft, Review and editing

Prof. Paul Flecknell
Roles: Review and editing

6.2 Additional scientific papers and edited book

Ormandy, E. *et al.*, Herrmann, K. (2019). Animal Research, Accountability, Openness and Public Engagement: Report from an International Expert Forum, *Animals,* 9, 622*.* Available at: https://www.mdpi.com/2076-2615/9/9/622/pdf

Herrmann, K., Pistollato, F. and Stephens, M. L. (2019). Beyond the 3Rs: Expanding the use of human-relevant replacement methods in biomedical research, *ALTEX - Alternatives to Animal Experimentation*, 36(3), pp. 343-352. Available at: https://www.altex.org/index.php/altex/article/view/1301

Herrmann, K. and Jayne, K. (2019). *Animal Experimentation: Working Towards a Paradigm Change.* Vol. 22. Leiden, The Netherlands: Brill. Available at: https://brill.com/view/title/35072?lang=en

Langley, G.R., Adcock, I.M., Busquet, F., Crofton, K.M., Csernok, E., Giese, C., Heinonen, T., Herrmann, K., Hofmann-Apitius, M., Landesmann, B. and Marshall, L.J. (2017). Towards a 21st-century roadmap for biomedical research and drug discovery: consensus report and recommendations. *Drug Discovery Today*, 22(2), pp. 327-339.

Davies *et al.*, Herrmann, K. (2016). Developing a collaborative agends for humanities and social scientific research on laboratory animal science and welfare. *PLoS ONE* 11(7), p. e0158791. Availabe at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158791

Herrmann, K. (2015). Tiermedizin. In: *Lexikon der Mensch-Tier-Beziehungen*, Hrsg. A. Ferrari und K. Petrus, Transcript, ISBN 978-3-8376-2232-4, pp. 352-354.
Translation: Explanation of the term “veterinary medicine” – history, present and future with focus on animal welfare in an encyclopedia on human-animal relationships.

Deckers, J. und Herrmann, K. (2012). Neue moralische Standards der EU Versuchstierrichtlinie – eine Chance für europäische Nutztiere auf Gleichbehandlung? *TIERethik,* 4. Jahrgang 2012/2, Heft 5, S. 87-94.
Translation: New moral standards of the EU Directive on the protection of animals used for scientific purposes (2010/63/U) **–** a chance for Europe's farmed animals to be treated equally?

Herrmann, K. (2012). Überwachung von Versuchstierhaltungen. *Amtstierärztlicher Dienst (ATD)*, 4/2012.
Translation: Federal inspections of laboratory animal husbandries.

Herrmann, K. (2011). In dubio pro animale – zum Ethikranking der bundesdeutschen Hochschulen*. TIERethik,* 3. Jahrgang 2011, Heft 3, S. 174-176.
Translation: In dubio pro animale – on the ethical ranking of German universities.

6.3 Contribution to conferences

6.3.1 Oral presentations

Herrmann, K. (2019). Refinement on the way towards replacement: Are we doing what we can?. *European Society for Alternatives to Animal Testing EUSAAT 2019 Congress*, Linz, Austria, October 2019. In: Alternatives to Animal Experimentation Proceedings, 8(1), p. 70. http://www.altex.ch/resources/altex_Linz2019_full.pdf

Herrmann, K. (2019). Full implementation of refinement methods in practice through education and training. *European Society for Alternatives to Animal Testing EUSAAT 2019 Congress*, Linz, Austria, October 2019. In: *Alternatives to Animal Experimentation Proceedings*, 8(1), p. 71. http://www.altex.ch/resources/altex_Linz2019_full.pdf

Herrmann, K. (2019). Education towards non-animal approaches in basic and applied biomedical research. *European Society for Alternatives to Animal Testing EUSAAT 2019 Congress*, Linz, Austria, October 2019. In: *Alternatives to Animal Experimentation Proceedings*, 8(1), p. 72. http://www.altex.ch/resources/altex_Linz2019_full.pdf

Herrmann, K. (2019). Education and Training to Fully Implement Refinement Methods in Practice. *70th American Association for Laboratory Animal Science (AALAS) National Meeting*, Denver, Colorado, USA, October 2019.

Herrmann, K. (2019). Animal Experimentation: Working Towards a Paradigm Change. *FRAME 50th Anniversary Symposium*, University of Nottingham, UK, July 2019. In: Alternatives to Laboratory Animals, 47(2), p. 95.
https://journals.sagepub.com/doi/pdf/10.1177/0261192919869383

Herrmann, K. (2019). Tierversuche: Auf dem Weg zu einem Paradigmenwechsel. *12. Tierversuchstagung „3R und Ersatzmethoden - bessere Forschung, weniger Tierleid"*, Olten, Switzerland, June 2019.

Herrmann, K. (2018). Animal Experimentation: Working Towards a Paradigm Change. *European Veterinary Congress for Behavioural Medicine and Animal Welfare* "Enhancing animal behaviour and welfare – through optimising human behaviour change", Berlin, September 2018. In: *Proceedings of the First Annual Meeting of the European Congress of Behavioural Medicine and Animal Welfare (ECVBMAW)*, pp. 5-9.

Herrmann, K. (2018). Towards a Paradigm Change: Biomedical animal experimentation and why public co-determination is important. *Research conference "New Directions in Animal Advocacy"*, University of Sydney, Australia, December 2018.

Herrmann, K. (2018). Translational failures call for critical appraisal of animal studies. *European Society for Alternatives to Animal Testing EUSAAT 2018 Congress*, Linz, Austria, September 2018. In: *Alternatives to Animal Experimentation Proceedings*, 7(2), p. 88.

Herrmann, K. (2018). Animal Experimentation: Harms and Flaws, and Ways to Work Towards a Paradigm Change. *4th Minding Animals International Conference (MAC4)*, Mexico City, Mexico, January 2018.

Herrmann, K. (2017). Refinement – what is in it for the animals?. *10th World Congress on Alternatives and Animal Use*, Seattle, USA, August 2017. In: *Alternatives to Animal Experimentation Proceedings*, 6(1), p. 13.

Herrmann, K. and Flecknell, P.A. (2016). The Culture of Care for Laboratory Animals in Germany with a focus on Refinement – Lessons learnt and recommendations for better practice. *European Congress on Animal Welfare & Behavioural Medicine (AWBM)*, Cascais, Portugal, October 2016.

Herrmann, K. (2016). Refinement von Tierexperimenten: Möglichkeiten und Grenzen der Minimierung von Schmerzen, Leiden und Angst von Ratten und Mäusen im Versuch. *Minding Animals Germany Symposium #3*, Erlangen, Germany, June 2016.
Translation: Refinement of animal experimentation: potentialities and limitations for the reduction of pain, suffering and anxiety of rats and mice used in experiments.

Herrmann, K. (2016). How and why is Public Opinion of Animal Experimentation important?. *Workshop on Animal Experimentation*, Institute for Advanced Study, Berlin, Germany, April 2016.

Herrmann, K. and Flecknell, P.A. (2015). A Critical Review of Animal Research Applications. *4th Annual EU Conference for Critical Animal Studies*, Lisbon, October 2015.

Herrmann, K. and Flecknell, P.A (2015). Pain Management and post-operative Monitoring and Care of Laboratory Rodents in Germany: Lessons learnt and Recommendations to improve practice. *European Society for Alternatives to Animal Testing EUSAAT 2015 Congress*, Linz, Austria, September 2015.

Herrmann, K. and Flecknell, P.A. (2015). A critical review of animal research applications. *Second Annual Summer School on Animal Ethics*, Oxford Centre for Animal Ethics, Oxford, UK, July 2015.

Herrmann, K. and Flecknell, P.A. (2015). The current state of Refinement in Germany. *CALAS/ACSAL 54th Annual Symposium*, Montreal, Canada, May/June 2015.

Herrmann, K. (2015). Animal experimentation: openness, public engagement and public concerns. *3rd Minding Animals International Conference (MAC3)*, New Dehli, January 2015.

Herrmann, K. and Flecknell, P.A. (2014). The Application of Refinement Methods in Germany. *Rus-LASA - ICLAS Conference*, St. Petersburg, Russia, October 2014.

Herrmann, K. and Flecknell, P.A. (2013). Assessment of the Efficacy of Anesthetic and Analgesic Regimen Involving Rodents in Germany, 64th AALAS National Meeting, Baltimore, USA, October 2013.

Herrmann, K. and Flecknell, P.A. (2013). A review of German animal research applications from 2010 to assess the appropriateness of the methods being used to kill laboratory rodents. *UFAW International Animal Welfare Science Symposium*, Barcelona, July 2013.

6.3.2 Poster presentations

Herrmann (2019). Teaching animal-free approaches in basic and applied biomedical research. *JRC Summer School on Non-Animal Approaches in Science – Challenges and Future Directions*, Ispra, Italy, May 2019.

Herrmann, K. and Flecknell, P.A. (2014). Severity classification in German animal research applications. *World Congress on Alternatives and Animal Use*, Prague, Czech Republic, August 2014.

Herrmann, K. and Flecknell, P.A. (2014). Refinement in German animal research applications. *World Congress* on *Alternatives* and Animal Use, Prague, Czech Republic, August 2014.

Herrmann, K. and Flecknell, P.A. (2013). An Assessment of the Efficacy of Anaesthetic and Analgesic Regimens involving Rodents in Germany. In: *12th FELASA 2013 SECAL Congress "Animal Research: Better Science with Fewer Animals".* Barcelona, June 2013. In: 12th FELASA SECAL Congress, *Journal of the American Association for Laboratory Animal Science*, 52(3), p. 387.

Herrmann, K. (2013). Directive 2010/63/EU – A Chance for More Humane Education?. In: *12th FELASA 2013 SECAL Congress "Animal Research: Better Science with Fewer Animals".* Barcelona, June 2013. In: 12th FELASA SECAL Congress, *Journal of the American Association for Laboratory Animal Science*, 52(3), p. 389.

Herrmann, K. and Flecknell, P.A. (2012). A review of German animal research applications from 2010 to identify which anaesthestic and analgesic regimens are used in experiments involving rodents. *LASA Winter Meeting*, Manchester, November 2012.

Herrmann, K. (2012). Directive (2010/63/EU) – an improvement for European laboratory animals?. *2nd Minding Animals Conference (MAC2)*, Utrecht, The Netherlands, July 2012

7 Acknowledgements

First of all, I would like to thank Heidrun Fink and Paul Flecknell for agreeing to act as my PhD supervisors. This work would have not been possible without their ongoing support. I am very grateful for Prof. Fink's courage in supporting a project that challenged the status quo, and for her help in applying for funding. And I am particularly appreciative of Paul's pioneering specialist expertise in experimental refinement methods which he generously shared with me. His mentoring has had a huge influence on my career.

I also want to thank the other members of Paul Flecknell's group who gave me an open-armed welcome when I came to Newcastle. I am thankful to Claire Richardson, the third person in my supervisory committee, for her mentorship during the time I spent at the Comparative Biology Center at Newcastle University's Medical School and beyond.

Besides Claire I had the pleasure of interacting with and learning from Huw Golledge, Johnny Roughan, Amy Miller, Yvette Ellen, Henri Bernand, Aurélie Thomas, Chris Blau and Matthew Leach.

My sincere thanks goes to Thorsten Busse, for his extensive assistance in developing the database that facilitated this research.

Turning the concept for this project into reality would have not been possible without the support of my former boss and head of Berlin's animal research inspectorate, Heidemarie Ratsch. She has been an important mentor to me since I began working with her in 2007. She not only helped me to acquire access to animal research proposals for this study and to apply for grants, but she also made it possible for me to reduce my work hours at the animal research inspectorate to dedicate time to this PhD project.

Furthermore, I thank Jörg Luy, Professor of animal welfare and ethics at my university until 2013, who kept on encouraging me to find a way to realize this first-of-its-kind review of animal research proposals.

I am thankful for the support of Torsten Nöldner, spokesman for animal welfare in the state of Berlin, who reached out to his colleagues at the other highest state authorities to ask for their participation in this study.

And I am grateful for Christoph Maisack, who in the name of the Deutsche Juristische Gesellschaft für Tierschutzrecht e.V. (German Association of Animal Welfare Law) has readily assisted me with any legal questions I have had over the years.

I gratefully acknowledge the funding sources that made my PhD work possible. This was first and foremost the German foundation SET, Stiftung zur Förderung von Ersatz- und Ergänzungsmethoden zur Einschränkung von Tierversuchen, who supported me for three years of the project, including my travel to various international conferences to present the project findings as well as my visits to Newcastle University. Thanks to ZEBET, the Center for Documentation and Evaluation of Alternative Methods to Animal Experiments, whose grant allowed me to travel to some of the competent authorities to retrieve research proposals for this study. Also, this extra grant gave me the opportunity to present my research results at additional conferences and meetings.

I am immensely grateful for all the dear friends who have given me their support on the, sometimes rocky, road towards completing my thesis. I especially want to thank Anne, Alka, Felix, Gilly, Jo, Lisa, Marty, and Stef.

I am very thankful for the encouragement of my beloved family: my mum who kept on nagging me to work faster and to focus on one project at a time, and my brother Klaus and his wife Anwen, who let me stay with them and took care of me during parts of the most intensive writing phase, when I was least fun to be around. And I am sincerely grateful to my late father; his passion for and dedication to the things he loved have been a constant inspiration.

And last but not least, I want to acknowledge the four-legged friends, who I had the pleasure to spend time with during some parts of the writing process: the dogs Lanai, Uri, Pender, Sparrow and Lucia, and the cats Dotty and Pebbles. They were best company, cheering me up when it was most needed, and frequently reminding me how fulfilling it is to contribute to making this world a better place for animals.

8 Funding sources

a. This study was funded by the German foundation SET, Stiftung zur Förderung von Ersatz- und Ergänzungsmethoden zur Einschränkung von Tierversuchen. SET fully supported this project for three years, including travel expenses to various international conferences to present the project findings as well as visits at Newcastle University.
ZEBET, the Center for Documentation and Evaluation of Alternative Methods to Animal Experiments, provided an additional grant (FK 1329-472) that allowed to travel to some of the competent authorities to retrieve research proposals for this study. Also, this extra grant gave the possibility to present the research results at additional conferences and meetings.

b. The funders had no influence on study design and conduct. The author declares that there is no conflict of interest.

9 Statement of authorship

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe.

Berlin, den 08.12.2019 Kathrin Herrmann