Check for updates

Original Article

Laboratory Animals Limited

Laboratory Animals
2026, Vol. 60(1) 110–118
© The Author(s) 2025

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00236772251374409
journals.sagepub.com/home/lan
Sage

One size does not fit all: Guidelines for designing a score sheet for animal experiments – eight essential steps

Michaela Thallmair[1], María J Duque-Correa[2], Maike Heimann[2] and Philippe Bugnon[2]

Abstract

The EU Directive 2010/63 on the protection of animals used for scientific purposes, as well as the Swiss Animal Welfare Legislation, demand monitoring and documentation of specific aspects of an animal experiment, including welfare-related issues and the (retrospective) assessment of the severity of the procedures that the animals underwent. A score sheet is an efficient tool for the evaluation of the burden of an animal during an experiment and, if properly designed and used, helps adhere to the 3Rs principle. It must be adapted to the specifics of each experiment and explicitly conceived for it. Several score sheet examples have been published; however, some contain fundamental flaws or are designed for specific settings only, requiring modifications to fit other experimental designs. This paper suggests an eight-step procedure to design a score sheet that can be adapted to any animal species and experimental conditions.

Keywords

Animal welfare, monitoring, severity assessment, score sheet, 3Rs

Date received: 17 April 2025; accepted: 8 August 2025

Introduction

Animal welfare legislations, such as the EU Directive 2010/63[1] and the Swiss Animal Welfare Law,[2] require monitoring, documentation and retrospective severity assessment in animal experiments. A key tool supporting these requirements is the score sheet, a structured form for consistent observation, standardized interventions, and structured evaluation of animal well-being.[3–5] It records clinical signs, guides decisions, and defines criteria for humane endpoints, helping to ensure both ethical compliance and scientific quality. Score sheets are based on the guidelines by Morton and Griffiths[6] and further developed by others.[4,7–9]

A score sheet must be tailored to the experimental setup to ensure trained personnel can perform a reproducible and standardized assessment. This enables timely interventions, reduces constraints, and ensures adherence to humane endpoints. Well-designed score sheets document the animal's burden and interventions to reduce the constraint. They also improve scientific validity, reliability, and data traceability.[4,10,11] Score sheets must match the species, interventions, and model used.

A literature evaluation on score sheets revealed that no publication explains how to design a score sheet (reviewed in Bugnon et al. 2016[4]). There are published sources on general guidance on score sheet use.[7,8,12–17] Recently, score sheets have been published for several species and specific research fields and models.[5,10,18–23] A collection of score sheets is accessible at https://journals.sagepub.com/page/lan/collections/score-sheets. Nevertheless, a previous analysis[4] showed that many published score sheets contain flaws that could lead to inadequate interventions and negatively affect animal welfare. Examples include incorrect sequencing of symptom monitoring, impairing proper assessment (e.g. weighing animals before observing undisturbed behaviour) or failing to cover certain ranges of body weight loss.[12,16,24–26] Regular reviews and updates to the score sheet are necessary, especially if novel findings emerge during monitoring.

[1]Office for Animal Welfare and 3R, University of Zurich, Switzerland
[2]Institute of Laboratory Animal Science, University of Zurich, Switzerland

Corresponding author:
Philippe Bugnon, Institute of Laboratory Animal Science, University of Zurich, Winterthurerstrasse 190, Zurich, 8057, Switzerland.
Email: philippe.bugnon@uzh.ch

This prevents missing important symptoms not listed on the score sheet and avoids diluting total scores by including parameters never observed in the model.[4,7,12,14,20,25] A thorough revision is necessary when refinements relevant to the study are published.

Without clear instructions, it is challenging to create a score sheet that detects relevant clinical symptoms, is easy to use, and leads to meaningful interventions and humane endpoints. We developed an interactive training to establish a score sheet scaffold adjustable for all animal experiments and species (https://vsfltkreg.uzh.ch/course/m-11/en). As a result, we devised an eight-step approach (Figure 1) to help design an efficient score sheet. Since May 2015, we have successfully demonstrated this method in our interactive continuing education courses.

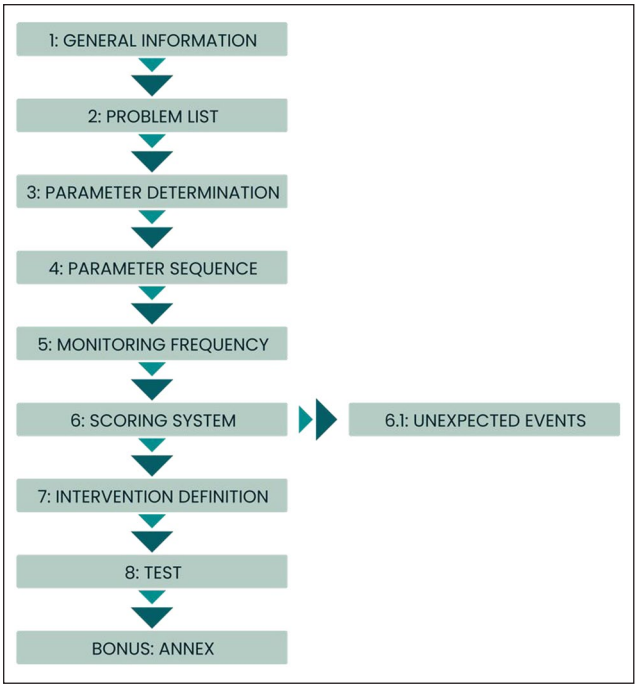


Figure 1. Flowchart outlining the steps to create a score sheet for animal experimentation.

Eight steps to create a score sheet for animal experiments

Step 1. General information

The score sheet should include key study information such as the study name, licence number, and the name and telephone number of the personnel to be contacted for questions or emergencies. It should also state the author and version number to ensure use of the latest version. Basic details include the animal identification number and group designation (e.g., control, treatment, or group code). Each score sheet must have space to record the date and time, if relevant (e.g., for multiple observations on the same day). At the bottom, space should be provided for the observer's name and signature. To ensure traceability for future use, such as interpreting outcomes, reporting the experiment, or preparing a manuscript, the general information allows the score sheet data to be clearly linked to a specific experiment and its corresponding animal or group.

Step 2. Problem list: which deviations from the animal's normal state are expected in the experimental setting?

This step is the most crucial part of creating the score sheet (Figure 1); however, it is often neglected when designing score sheets. This step aims to develop an extensive list of possible problems and deviations from the normal state, based on the planned experimental interventions. It is necessary to consider general as well as experiment- or model-specific deviations from the animal's normal state.[6,27,28] The problem list should comprise behavioural and physiological deviations[14] as listed in Table 1. Familiarity with the experimental model, the species in question, and the behavioural repertoire of the species is a prerequisite.[29] Furthermore, a thorough literature search is required to gain important information regarding what has been reported in similar experimental conditions. The experimenters should also consider other parameters that are not directly related to the experimental conditions. The animal's intrinsic factors, such as sex, age, strain, or genotype, can affect its health and well-being.[30–33] This further emphasizes the importance of having deep knowledge regarding the animals and ensures that the list of problems corresponds to the experimental setting. The expected deviations/problems recognized in this step are not the parameters to be assessed and listed on the score sheet, but they will serve to prepare the ground for these parameters.

Step 3. Parameter determination: which indicators help to detect deviations from the normal state?

Building on the list created during Step 2, it is helpful to formulate a series of questions that will support the determination of the parameters to detect deviations from the normal state (Table 1). Ideally, the parameters are robust and deliver comparable results between different raters or laboratories. For every problem detected in the previous step, researchers should consider all possible options to identify and subsequently evaluate the problem.[14] It is crucial at this stage that experimenters determine which parameters are most appropriate for their expertise and possibilities (Figure 2). Similarly, all chosen parameters should be sensitive and specific, facilitating the detection of the intended deviation with a high success rate.[34] The chosen parameters must be suited to the animal species, the experimental

Table 1. Examples of deviations from normal state (Step 2) and possible questions that will help to define parameters to assess these deviations (Step 3). Note that this list is non-exhaustive, and not all parameters are suited for all species and all experimental settings.

Deviation category	Step 2: list of possible problems/deviations	Step 3: questions that will define the parameters to detect the problems/ deviations
Behavioural	Grooming	Is grooming behaviour changed (e.g., order or more attention to a specific body part)? How is the fur condition? Is there evidence of porphyrin staining?
	Social behaviour	Has the interaction with other animals changed? Is the animal (being) isolated? Does it show apathy? Has a once calm and docile animal suddenly become anxious or aggressive?
Physiological	Pain	Does the animal show any pain signs (e.g. hunchback, . . . or, e.g., for abdominal pain: does the animal show a firm belly, is the abdomen pressed to the ground?) Does the animal show attention to a certain body part (e.g. more licking, looking and/or grooming)?
	Water intake/hydration	How much water is consumed? Is the skin turgor (skin pinch test) normal?
	Food intake	Are there changes in the body weight or food intake? Are there changes in body condition score?
	Tumour growth	Is there an evident (palpable) tumour? Has the tumour size changed? Any necrosis or ulceration?
	Gastrointestinal problems	Has the faecal output changed? Is the faecal consistency or colour different?

conditions, and the experience of the people responsible for monitoring the animals and using the score sheet. Parameter choice can also be guided by published sources. Recently, there have been some efforts to publish composite scores for specific experimental conditions.[22,23]

Some deviations are easier to assess than others. For example, signs of pain and early infection signs are not always easy to recognize. Thus, experimenters must carefully consider which indicators might be helpful for monitoring. Furthermore, some parameters are only useful under specific experimental conditions. For example, reduced water and food intake can be easily identified when animals are single-housed; however, the interpretation of this parameter in group-housed animals is limited since such values give no information on an individual animal.[35] Similarly, some score sheets suggest using heart or breathing rate as parameters. These indicators are very helpful for assessing the animal's state but are difficult or impossible to measure in small species, such as rodents or aquatic species.

For this step, knowing the expected behaviour and physiology of the species and the individual animal is a prerequisite.[29,36] When prey animals are in pain, they may exhibit 'normal' behaviour as an innate response, since this may protect them from predators; this is known as 'displacement behaviour'.[35] Furthermore, stress due to manipulation can affect the animal's behaviour (see Step 4). For example, it is difficult to observe small rodents in their cages without moving or opening the cages; therefore, observed behaviours will be altered.[35] Similarly, if an animal has undergone surgery and cannot move properly owing to the procedure, this must be considered when assessing the animal. The same applies when analgesia is given; depending on the drug, behavioural changes such as apathy, reduced food intake, or increased activity can be observed.[37] It should further be noted that the observer can affect the behaviour of the observed animals. Specifically, it has been shown that rodents respond differently to male and female observers and experimenters, and handling by male personnel induces increased stress.[38] It has been proposed that the stress response related to males can suppress pain behaviours.[39] Similarly, it has been demonstrated that rodents respond to the affective state of the handler.[40] Therefore, during the interpretation of the scoresheet data, intrinsic factors related to the observer should be considered. Finally, the time of day when monitoring is done has to be carefully considered since it might influence the parameters to be observed. Furthermore, the light or dark phase is directly related to the activity levels of the animals; if observations are done during the inactive phase, behaviours will likely be missed (see Step 4).

Step 4. Parameter sequence

All manipulations to the cage or the animal itself may change the animal's behaviour or stress level, especially when working with prey animals.[35] The order in which parameters are assessed is, therefore, of great importance.[7] This should be reflected in the design of the score sheet.

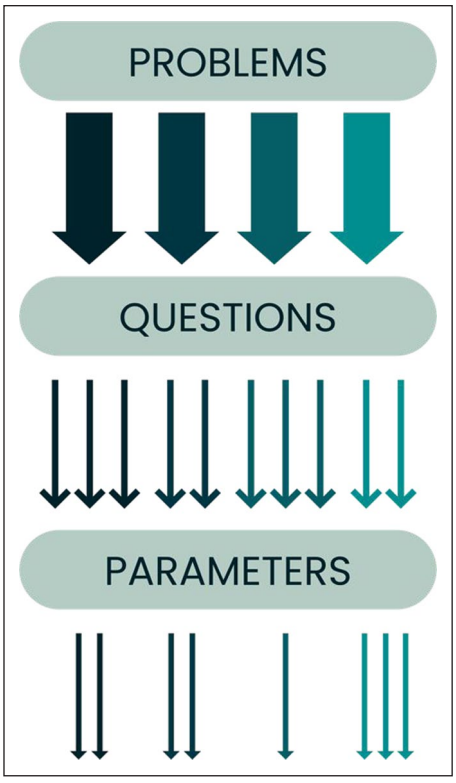


Figure 2. Flowchart showing the progression from Step 2 to Step 3. Once the putative problem list has been created, coming up with several questions per problem will help in a later step to decide which parameters can be used to monitor the individual animal. Each colour represents a different problem. Note that there is variability in the number of questions and parameters per problem.

The score sheet is intended to be read from top to bottom, like a text or list. However, if the parameter order is not adapted to this 'reading scheme' it might interfere with the assessment of symptoms, making it impossible to observe certain parameters/behaviours, or bias the result of the measurements and observations. Therefore, in this step, the list of parameters from Step 3 will be ordered from pure observation without any stimulus to observations and tests that would, when performed first, interfere with the next measurements. Similarly, some parameters can be influenced by stress (i.e. heart rate), whereas others remain unchanged (i.e. body mass). Therefore, it is better to start with visual observations that do not require any manipulation (e.g. assessment of social behaviour or posture) followed by responses to stimuli (e.g. provoked behaviour) to those that involve manipulation (e.g. heart rate, body temperature measurement, body weight) (Figure 3).

Step 5. Monitoring frequency

The frequency of observations should be adapted to the experimental setting. It must be frequent enough to allow for the timely detection of changes in the animals' well-being.[7] It is important to adjust the evaluation frequency to the different experimental phases, as the monitoring requirements may vary considerably. For example, in a cancer study involving subcutaneous tumours, the animal's state will not be impacted at the start of the experiment when the tumour is small and not metastasized. In this early phase, the monitoring frequency may be lower, to not burden the animal with unnecessary scoring stress. However, as the tumour grows or metastasizes, the frequency must be increased, as negative effects on the animal are to be expected. The opposite applies to the monitoring frequency after surgery. In this case, more frequent observation is required at the beginning (immediately after the procedure), as pain and stress occur immediately after surgery and, therefore, require close observation to assess the efficacy of the analgesic treatment. However, as adverse effects subside over time and analgesia is administered, longer intervals between observations are possible if no adverse events are observed. The researchers should reflect on the experimental design and determine when to adjust the observation frequency. For animal welfare, it is best to determine the observation frequency based on the cumulative score of the animals, increasing the frequency of monitoring if welfare is impaired.

Additionally, as discussed in Step 3, the effect of time of day on observations should be considered.[14] It is advisable to observe the animals during their active phase, that is, for most rodents, during the dark phase, to ensure accurate monitoring of all parameters and to minimize the likelihood of missing important indicators of declining well-being. If monitoring during the animals' active phase is not possible, the choice of parameters should reflect the fact that the animals are checked during their non-active phase. Furthermore, to increase the reliability of the score sheet, the observations should be done at the same time every day.

Step 6. Scoring system and grading the severity of the signs

To develop a system to grade the severity of the observed changes, the parameters defined in Step 3 are used (Figure 1 and Table 2). This allows for a systematic assessment of the animal's state, as well as continuous monitoring of its welfare, since scores can be compared between successive monitoring points.[14]

The scoring system can be either a simple binary system (yes/no, present/absent), or a numerical score that weighs/quantifies the symptoms and can account for cumulative harm by summing the individual scores.[14,41] A combination of both systems is also possible. Each system, whether binary or numerical, has relative strengths and weaknesses; some publications provide tables to guide the researchers in choosing the best system (or system combination) in a

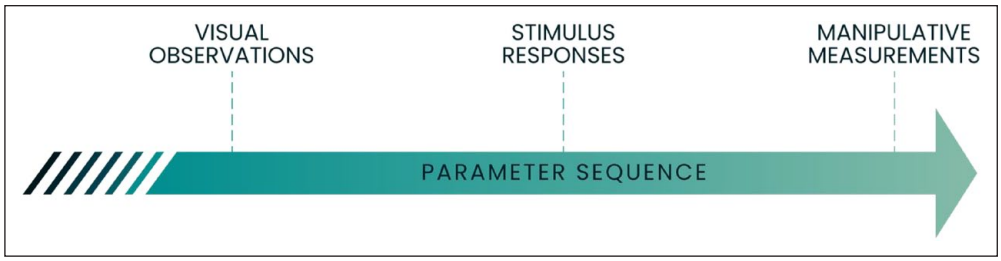


Figure 3. Parameter sequence. Once all parameters have been decided, they should be ordered in a logical sequence. Parameters that require only observation are assessed first, response to stimulus is tested afterwards and the final round includes recording parameters that require manipulation.

Table 2. Example of binary and numerical systems within a score sheet.

Parameter 1		Parameter 2		Parameter 3		Parameter 4		Parameter 5		Parameter 6	
Yes/absent	0	Severity	0	Severity	0	Severity	0	Severity	0	Severity	0
No/present	1	Severity	1	Severity	1	Severity	1	Severity	1	Severity	1
		Severity	2	Severity	3	Severity	2	Severity	3	Severity	7
		Severity	3	Severity	5	Severity	5	Severity	7	Severity	9

specific context.[14,15,42,43] A binary score can be used for parameters whose intensity is difficult or impossible to assess; only presence or absence of the parameter is noted (Table 2). Binary scores are also useful for parameters that lead to immediate decisions, such as termination criteria (humane endpoints). Numerical scores aim at quantifying the severity of signs or parameters. A score of '0' means normal animal state, and increasing values indicate increasing severity and impairment of the animal's state.[14,20,25] Additionally, the use of a non-regular scoring system (e.g. 0–1–5–6) is possible, if necessary, as each score given to a situation reflects the quantification of the constraint (Table 2). If the researcher believes that one of the parameters has a greater impact on the animal's health and well-being, a higher score could be assigned, accounting for the greater weight of a certain symptom.[14]

Regardless of the scoring system, calculating a cumulative score by summing the individual parameter scores should be considered.[13,14] This cumulative score reflects the overall burden on the animal at a given time point and supports decisions regarding observation frequency and predefined termination criteria. However, interventions should be based on single parameters, addressing each problem individually. When defined scores are reached, either for the individual parameters or the cumulative scores, humane intervention and endpoints, respectively, should be defined (see Step 7).[9] Furthermore, existing quantitative measurements, such as body condition score,[44,45] can be included as scoring systems for specific parameters.

This step is completed by adding a designated space to record the actual severity degree based on the cumulative scores of all parameters assessed. If the severity degree of the experimental animals needs to be reported to the authorities this information can then be easily collected and provided based on the data of the individual animals.[8]

Step 6.1. Unexpected events or nothing abnormal detected. The score sheet should be simple when an animal's well-being appears normal. Thus, including a checkbox such as 'No abnormality detected' can streamline documentation. This feature facilitates efficient use of the score sheet while still ensuring that both the monitoring process and the animal's normal state are properly recorded. On the other hand, symptoms might arise in an animal experiment that were not expected. Therefore, space to record detailed information of such unexpected events is needed. Recording such information is important, as it may impact the experiment and, consequently, the interpretation of the results. If such unexpected findings are recorded more often, they should be included in the score sheet.

Step 7. Definition of the interventions

The experimenter should always strive to minimize the animal's burden as much as possible within the context of the study. In this step, the researcher determines when actions or humane interventions should be taken and when humane endpoints are reached (according to Step 6).[46] When a deviation from an animal's normal state is identified, predefined humane intervention points must be established to prevent further deterioration of the animal's state.[9] For each scenario outlined in Step 3, suitable interventions should be defined, with clearly described procedures to ensure that appropriate measures can be implemented promptly and effectively. Such interventions include, but are not limited to, fluid replacement for dehydration, analgesic treatment for pain,

and providing easily accessible food for animals with mobility impairments or weakness.[47] The score sheet must also include re-evaluations to assess the success of the defined interventions (e.g., to assess the effectiveness of painkillers). If interventions were not successful, further actions must be predefined, such as additional interventions or experiment termination.[9] The humane interventions or termination procedures are to be included on the score sheet.

In some experimental settings, such as infection models, deviations are expected and do not grant interventions for the sake of the experiment; however, for animal welfare and for scientific reasons, monitoring should always be done, and all deviations noted.

Step 8. Testing the score sheet ('play the game')

The objective of this step is to identify any potential issues that may have been missed in previous steps and to make adjustments before the experimental phase begins. At this stage, all personnel involved should participate in the evaluation, ideally by rehearsing scenarios representing different levels of change in the animals' state. This step also ensures that everyone understands how to use the score sheet. In countries where the score sheet is a mandatory part of the experimentation licence, it helps reduce the need for amendments that might arise if issues are noticed later.

We recommend reviewing each step carefully so that all deviations from the normal state are considered and few modifications are needed during the study. Nevertheless, changes to the score sheet should be made whenever necessary, and the version updated accordingly. If the score sheet is part of a licence, any new version must be approved by the authorities.

Bonus Step. Score sheet annex

As a useful addition, we recommend creating an annex to the score sheet. This annex contains supporting documents that are not required for daily evaluations but help ensure quick and accurate responses when intervention is needed. For example, in experiments where skin injuries or tumours are expected, a diagram of the animal (dorsal, ventral, lateral views) allows for drawing the location and size of lesions, making it easier to monitor progress. The annex can also include detailed procedures for interventions (Step 7), humane euthanasia protocols, and instructions for sample collection and processing in the case of premature termination. These documents ensure that important samples are not lost and support reproducibility. Any other relevant materials may also be included. The annex should be stored with the score sheet for easy access in the event of unexpected situations.

Discussion

The development and implementation of score sheets in animal experimentation are key to refining procedures, enhancing animal welfare, and ensuring scientific validity. Legal frameworks in Europe and Switzerland require monitoring and retrospective severity assessment, but do not prescribe specific methods. Score sheets are the most suitable tool to meet these requirements. Other monitoring options exist, including endpointR,[48] and RELSA.[49] However, software such as endpointR only accounts for one variable, whilst well-being deterioration is multifactorial, and it is possible that using such an approach leads to under- or over-interpretation of the animal's state. On the other hand, a composite measures scale like RELSA is more comprehensive, while lacking validation, and it also does not include behavioural parameters. Therefore, until further development of other monitoring tools, score sheets are the most suitable tool to meet the legal requirements.

Our eight-step approach offers a structured, broadly applicable framework that enhances ethical responsibility and experimental rigour.

One size does not fit all in monitoring animal well-being. Species, interventions, and housing conditions vary, requiring individualized tools. Literature provides many examples, but few offer systematic guidance on designing score sheets effectively. Existing sheets often lack transparency in parameter selection, order, and scoring logic, leading to compromised assessments and suboptimal care. Our method emphasizes proactive planning and adaptation, starting with analysis of deviations from normal behaviour and physiology.

Our framework promotes flexibility and regular review of parameters and thresholds based on new data or unforeseen events. Score sheets are living documents evolving through observation. Including steps such as testing and maintaining an annex ensures that both routine and exceptional situations are addressed efficiently. Involving all personnel in score sheet development and rehearsal fosters shared understanding, consistent observation and quicker responses. This supports broader efforts in lab animal science to improve training and accountability.

A well-designed score sheet enables the early detection of adverse effects, allowing for intervention before serious harm. It supports the 3Rs principle, particularly refinement, by encouraging evidence-based and timely interventions. This strengthens both animal welfare and the outcomes' reliability, crucial for scientific integrity and publication.

While grounded in practical experience and scientific review, our eight-step guide remains adaptable. New technologies, such as automated monitoring or AI-based behaviour analysis, might enhance the monitoring methodology. Berce (2024)[5] evaluated AI-generated score sheets and concluded that thorough review and editing are necessary.

Future research should explore how to integrate such tools without compromising core ethical and scientific standards.

In conclusion, our framework enables researchers to develop tailored, and scientifically robust score sheets, bridging regulations, ethics and precision to improve both animal care and research quality.

Conclusion

A carefully constructed score sheet is key to promoting refinement in animal experimentation, contributing to improved animal welfare and enhanced scientific quality. As a living document, it should be reviewed and adjusted during the study in response to observed signs and unexpected findings. Systematic documentation of clinical observations and interventions ensures compliance with regulatory and ethical standards and supports more robust data interpretation and reproducibility. The eight-step framework presented here offers a flexible, scientifically grounded method for developing score sheets tailored to specific experimental models and species, providing researchers with a reliable means to detect relevant changes in well-being and implement timely, appropriate interventions.

Acknowledgements

We are grateful to Catalina Martinez-Guerra for her valuable assistance in designing all figures included in this manuscript. We also thank Prof. Dr Thorsten Buch and Prof. Dr Christopher Pryce for their thoughtful revisions of the text.

Data availability statement

This study did not involve the collection, generation or analysis of any quantitative or qualitative datasets. The research is based entirely on conceptual, theoretical or methodological work, and does not rely on empirical data. Therefore, there are no data available for sharing or archiving. As a result, no data repository links or contact information for data access are provided.

Declaration of conflicting interests

The authors have no conflicts of interest to declare.

Ethics statement

This project did not require an ethical board approval because it did not involve human or animal trials.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Michaela Thallmair https://orcid.org/0000-0002-9211-5643
María J Duque-Correa https://orcid.org/0000-0001-8431-2228
Maike Heimann https://orcid.org/0000-0002-5052-8757
Philippe Bugnon https://orcid.org/0000-0003-0680-2808

References

1. European Parliament, Council of the European Union. 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. In: *Official Journal of the European Union*, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:en:PDF (2010, accessed 13 March 2025).
2. Federal Assembly of the Swiss Confederation. Animal Welfare Act of 16 December 2005, https://www.fedlex.admin.ch/eli/cc/2008/414/en (2005, accessed 12 September 2024).
3. Golledge H and Jirkof P. Score sheets and analgesia. *Lab Anim* 2016; 50: 411–413.
4. Bugnon P, Heimann M and Thallmair M. What the literature tells us about score sheet design. *Lab Anim* 2016; 50: 414–417.
5. Berce C. Artificial intelligence generated clinical score sheets: looking at the two faces of Janus. *Lab Anim Res* 2024; 40: 21.
6. Morton DB and Griffiths PH. Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *Vet Rec* 1985; 116: 431–436.
7. Morton DB. The importance of non-statistical design in refining animal experiments. *REDVET Revista electrónica de Veterinaria* 2008; 9: 1–14.
8. Smith D, Anderson D, Degryse AD, et al. Classification and reporting of severity experienced by animals used in scientific procedures: FELASA/ECLAM/ESLAV Working Group report. *Lab Anim* 2018; 52: 5–57.
9. Williams WO and Baneux P. Humane intervention points: Refining endpoint terminology to incorporate non-euthanasia intervention options to improve animal welfare and preserve experimental outcomes. *Lab Anim* 2022; 56: 482–489.
10. Nelson K, McFall A, Stonecypher T, et al. Development and validation of a scoring system for abnormalities in the Gopher Frog (*Rana capito*). Savannah River Ecology Laboratory, https://www.osti.gov/biblio/2370039 (2024, accessed 14 July 2025).
11. Meagher RK. Observer ratings: Validity and value as a tool for animal welfare research. *Appl Anim Behav Sci* 2009; 119: 1–14.
12. Kendall LV, Owiny JR, Dohm ED, et al. Replacement, refinement, and reduction in animal studies with biohazardous agents. *ILAR J* 2018; 59: 177–194.
13. Ullmann K, Jourdan T, Kock M, et al. Recommendations for the development and use of score sheets as a tool for applied refinement, https://www.cabidigitallibrary.org/doi/full/10.5555/20183235029 (2018, accessed 12 September 2024).
14. Hawkins P, Morton DB, Burman O, et al. A guide to defining and implementing protocols for the welfare assessment of laboratory animals: Eleventh report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Lab Anim* 2011; 45: 1–13.
15. Morton DB and Hau J. Welfare assessment and humane endpoints. In: Hau J and Schapiro S J. (eds) *Handbook of laboratory animal science*. Boca Raton: CRC Press, 2002.
16. Lloyd MH and Wolfensohn SE. Practical use of distress scoring systems in the application of humane endpoints. *Humane*

endpoints in animal experiments for biomedical research. 1999;48–53.
17. Wolfensohn S and Lloyd M. Pain, stress and humane end points. In: *Handbook of laboratory animal management and welfare* 2003, pp.59–73.
18. Koch A, Gulani J, King G, et al. Establishment of early endpoints in mouse total-body irradiation model. *PLoS One* 2016; 11: e0161079.
19. Izer JM, LaFleur RA, Weiss WJ, et al. Development of a pain scoring system for use in sheep surgically implanted with ventricular assist devices. *J Invest Surg* 2019; 32: 706–715.
20. Hagemeister K, Ernst L, Kadaba Srinivasan P, et al. Severity assessment in pigs after partial liver resection: Evaluation of a score sheet. *Lab Anim* 2020; 54: 251–260.
21. Flueck-Giraud M, Schmidt-Posthaus H, Bergadano A, et al. An adaptable, user-friendly score sheet to monitor welfare in experimental fish. *Lab Anim* 2025; 59: 226-241.
22. Talbot SR, Kumstel S, Schulz B, et al. Robustness of a multivariate composite score when evaluating distress of animal models for gastrointestinal diseases. *Sci Rep* 2023; 13: 2605.
23. Van Dijk RM, Koska I, Bleich A, et al. Design of composite measure schemes for comparative severity assessment in animal-based neuroscience research: A case study focussed on rat epilepsy models. *PLoS One* 2020; 15: e0230141.
24. Huet O, Ramsey D, Miljavec S, et al. Ensuring animal welfare while meeting scientific aims using a murine pneumonia model of septic shock. *Shock* 2013; 39: 488–494.
25. Fentener Van Vlissingen J, Borrens M, Girod A, et al. The reporting of clinical signs in laboratory animals: FELASA Working Group Report. *Lab Anim* 2015; 49: 267–283.
26. Baier J, Rix A, Drude NI, et al. Influence of MRI examinations on animal welfare and study results. *Invest Radiol* 2020; 55: 507–514.
27. Drude N, Pawlowsky K, Tanaka H, et al. Severity assessment in rabbits after partial hepatectomy: Part II. *Lab Anim* 2016; 50: 468–475.
28. Graf R, Cinelli P and Arras M. Morbidity scoring after abdominal surgery. *Lab Anim* 2016; 50: 453–458.
29. Carstens E and Moberg GP. Recognizing pain and distress in laboratory animals. *ILAR J* 2000; 41: 62–71.
30. Zhang H, Lin S, Chen X, et al. The effect of age, sex and strains on the performance and outcome in animal models of stroke. *Neurochem Int* 2019; 127: 2–11.
31. Jackson SJ, Andrews N, Ball D, et al. Does age matter? The impact of rodent age on study outcomes. *Lab Anim* 2017; 51: 160–169.
32. Gilbert MJ, Zerulla TC and Tierney KB. Zebrafish (*Danio rerio*) as a model for the study of aging and exercise: Physical ability and trainability decrease with age. *Exp Gerontol* 2014; 50: 106–113.
33. Van den Bos R, Flik G and Gorissen M. Behavioral research in zebrafish (*Danio rerio*): Strain as source of variation. In: Gerlai RT (ed.) *Behavioral and neural genetics of zebrafish*. London: Academic Press, 2020, pp.245–262.
34. Jirkof P, Abdelrahman A, Bleich A, et al. A safe bet? Inter-laboratory variability in behaviour-based severity assessment. *Lab Anim* 2020; 54: 73–82.
35. Kohn DF, Martin TE, Foley PL, et al. Guidelines for the assessment and management of pain in rodents and rabbits. *J Am Assoc Lab Anim Sci* 2007; 46: 97–108.
36. Stasiak KL, Maul DON, French E, et al. Species-specific assessment of pain in laboratory animals. *J Am Assoc Lab Anim Sci* 2003; 42: 13–20.
37. Jeger V, Hauffe T, Nicholls-Vuille F, et al. Analgesia in clinically relevant rodent models of sepsis. *Lab Anim* 2016; 50: 418–426.
38. Sorge RE, Martin LJ, Isbester KA, et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat Methods* 2014; 11: 629–632.
39. Faraji J, Ambeskovic M, Sauter N, et al. Sex-specific stress and biobehavioral responses to human experimenters in rats. *Front Neurosci* 2022; 16: 965500.
40. Kaźmierowska AM, Kostecki M, Szczepanik M, et al. Rats respond to aversive emotional arousal of human handlers with the activation of the basolateral and central amygdala. *Proc Natl Acad Sci U S A* 2023; 120: e2302655120.
41. Honess P and Wolfensohn S. The Extended Welfare Assessment Grid: A matrix for the assessment of welfare and cumulative suffering in experimental animals. *Altern Lab Anim* 2010; 38: 205–212.
42. Morton DB. Humane endpoints in animal experimentation for biomedical research: Ethical, legal and practical aspects. In: *Humane endpoints in animal experiments for biomedical research*. London: Royal Society of Medicine Press, 1998, pp.5–12.
43. Hendriksen C, Morton D and Cussler K. Use of humane endpoints to minimise suffering. In: *The COST manual of laboratory animal care and use*. Boca Raton: CRC Press, 2016, pp.333–353.
44. Ullman-Culleré MH and Foltz CJ. Body condition scoring: A rapid and accurate method for assessing health status in mice. *Comp Med* 1999; 49: 319–323.
45. Hickman DL and Swan M. Use of a body condition score technique to assess health status in a rat model of polycystic kidney disease. *J Am Assoc Lab Anim Sci* 2010; 49: 155–159.
46. Morton DB. A systematic approach for establishing humane endpoints. *ILAR J* 2000; 41: 80–86.
47. Hrapkiewicz K, Colby LA and Denison P. *Clinical laboratory animal medicine: An introduction*. Iowa: John Wiley & Sons, 2013.
48. Helgers SO, Talbot SR, Riedesel AK, et al. Body weight algorithm predicts humane endpoint in an intracranial rat glioma model. *Sci Rep* 2020; 10: 9020.
49. Talbot SR, Struve B, Wassermann L, et al. RELSA—A multidimensional procedure for the comparative assessment of well-being and the quantitative determination of severity in experimental procedures. *Front Vet Sci* 2022; 9: 937711.

Une taille unique ne convient pas à tous : lignes directrices pour la conception d'une feuille de notation des expériences animales – huit étapes essentielles

Résumé

La directive 2010/63 de l'UE sur la protection des animaux utilisés à des fins scientifiques, ainsi que la législation suisse sur le bien-être des animaux, exigent le suivi et la documentation d'aspects spécifiques d'une expérience animale, notamment les questions liées au bien-être et l'évaluation (rétrospective) de la gravité des procédures subies par les animaux. Une feuille de notation constitue un outil efficace pour évaluer le fardeau d'un animal au cours d'une expérience et, si elle est correctement conçue et utilisée, elle permet de respecter le principe des 3R. Elle doit être adaptée aux spécificités de chaque expérience et explicitement conçue à cette fin. Plusieurs exemples de feuilles de notation ont été publiés; certaines contiennent toutefois des défauts fondamentaux ou sont conçues uniquement pour des paramètres spécifiques, nécessitant donc des modifications pour s'adapter à d'autres conceptions expérimentales. Cet article propose une procédure en huit étapes pour concevoir une feuille de notation susceptible d'être adaptée à toutes les espèces animales et conditions expérimentales.

Ein Modell für alle gibt es nicht: Leitlinien für die Gestaltung eines Bewertungsbogens für Tierversuche – acht wesentliche Schritte

Abstract

Die EU-Richtlinie 2010/63 zum Schutz von Tieren, die für wissenschaftliche Zwecke verwendet werden, sowie die Schweizer Tierschutzgesetzgebung verlangen die Überwachung und Dokumentation bestimmter Aspekte eines Tierversuchs, darunter Tierschutzaspekte und die (retrospektive) Bewertung des Schweregrads der Verfahren, denen die Tiere unterzogen wurden. Ein Bewertungsbogen ist ein effizientes Instrument zur Bewertung der Belastung eines Tieres während eines Experiments und trägt, wenn er richtig gestaltet und verwendet wird, zur Einhaltung des 3R-Prinzips bei. Er muss an die Besonderheiten jedes Versuchs angepasst und ausdrücklich dafür konzipiert sein. Es wurden mehrere Beispiele für Bewertungsbögen veröffentlicht, jedoch weisen einige grundlegende Mängel auf oder sind nur für bestimmte Settings konzipiert, sodass sie für andere Versuchsdesigns entsprechend angepasst werden müssen. In diesem Artikel wird ein achtstufiges Verfahren zur Erstellung eines Bewertungsbogens vorgeschlagen, der an beliebige Tierarten und beliebige Versuchsbedingungen angepasst werden kann.

Una sola fórmula no sirve para todos: directrices para diseñar una hoja de puntuación en experimentos con animales – ocho pasos esenciales

Resumen

La Directiva 2010/63 de la UE sobre la protección de los animales utilizados con fines científicos, así como la Legislación Suiza de Bienestar Animal, exigen el seguimiento y la documentación de aspectos específicos de un experimento animal, incluidos los relacionados con el bienestar y la evaluación (retrospectiva) de la severidad de los procedimientos a los que fueron sometidos los animales. Una hoja de puntuación es una herramienta eficaz para la evaluación de la carga de un animal durante un experimento y, si se diseña y utiliza correctamente, ayuda a cumplir el principio de las 3 erres. Debe adaptarse a las particularidades de cada experimento y concebirse de forma explícita para este. Se han publicado varios ejemplos de hojas de puntuación; sin embargo, algunas contienen defectos fundamentales o están diseñadas únicamente para entornos específicos, por lo que requieren modificaciones para adaptarse a otros diseños experimentales. Este artículo sugiere un procedimiento de ocho pasos para diseñar una hoja de puntuación que pueda adaptarse a cualquier especie animal y condiciones experimentales.