Video Article

Systematic Assessment of Well-Being in Mice for Procedures Using General Anesthesia

Katharina Hohlbaum[1,2], Bettina Bert[2,3], Silke Dietze[2], Rupert Palme[4], Heidrun Fink[2], Christa Thöne-Reineke[1]

[1]Institute of Animal Welfare, Animal Behavior and Laboratory Animal Science, Department of Veterinary Medicine, Freie Universität Berlin

[2]Institute of Pharmacology and Toxicology, Department of Veterinary Medicine, Freie Universität Berlin

[3]German Federal Institute for Risk Assessment (BfR)

[4]Unit of Physiology, Pathophysiology and Experimental Endocrinology, Department of Biomedical Sciences, University of Veterinary Medicine

Correspondence to: Katharina Hohlbaum at Katharina.Hohlbaum@fu-berlin.de

URL: https://www.jove.com/video/57046
DOI: doi:10.3791/57046

Keywords: Behavior, Issue 133, welfare, well-being, distress, severity classification, severity assessment, fecal glucocorticoid metabolites, mice, refinement, behavioral test, luxury behavior, 3R, anesthesia, Mouse Grimace Scale

Date Published: 3/20/2018

Citation: Hohlbaum, K., Bert, B., Dietze, S., Palme, R., Fink, H., Thöne-Reineke, C. Systematic Assessment of Well-Being in Mice for Procedures Using General Anesthesia. *J. Vis. Exp.* (133), e57046, doi:10.3791/57046 (2018).

Abstract

In keeping with the 3R Principle (Replacement, Reduction, Refinement) developed by Russel and Burch, scientific research should use alternatives to animal experimentation whenever possible. When there is no alternative to animal experimentation, the total number of laboratory animals used should be the minimum needed to obtain valuable data. Moreover, appropriate refinement measures should be applied to minimize pain, suffering, and distress accompanying the experimental procedure. The categories used to classify the degree of pain, suffering, and distress are non-recovery, mild, moderate, or severe (EU Directive 2010/63). To determine which categories apply in individual cases, it is crucial to use scientifically sound tools.

The well-being-assessment protocol presented here is designed for procedures during which general anesthesia is used. The protocol focuses on home cage activity, the Mouse Grimace Scale, and luxury behaviors such as burrowing and nest building behavior as indicators of well-being. It also uses the free exploratory paradigm for trait anxiety-related behavior. Fecal corticosterone metabolites as indicators of acute stress are measured over the 24-h post-anesthetic period.

The protocol provides scientifically solid information on the well-being of mice following general anesthesia. Due to its simplicity, the protocol can easily be adapted and integrated in a planned study. Although it does not provide a scale to classify distress in categories according to the EU Directive 2010/63, it can help researchers estimate the degree of severity of a procedure using scientifically sound data. It provides a way to improve the assessment of well-being in a scientific, animal-centered manner.

Video Link

The video component of this article can be found at https://www.jove.com/video/57046/

Introduction

EU Directive 2010/63[1] stipulates that the 3R Principle (Replacement, Reduction, Refinement) developed by Russel and Burch[2] is to be applied whenever animal experimentation is necessary. The ultimate goal of the EU Directive is to phase out all animal testing, but the Directive acknowledges that, for the time being, some animal experiments are still needed to conduct research that will protect human and animal health. Thus, if an animal experiment cannot be replaced by any alternative method, only the minimum number of laboratory animals is to be used to obtain reliable results. In addition, the amount of pain, suffering, and distress accompanying experimental procedures should be minimized using appropriate refinement measures. EU Directive 2010/63 stipulates that the severity of a procedure must be prospectively classified as non-recovery, mild, moderate, or severe[1]. As severity classification is decided on a case-by-case basis, it is important to have scientifically sound tools to estimate the severity of a given procedure.

Score sheets as proposed by Morton and Griffith[3] are an essential tool in detecting any deviations from normal status, including negative effects on well-being[4]. Score sheets are used to retrospectively determine pain, suffering, and distress caused by an experiment and focus on visible changes in the physical state of the individual animal (*e.g.*, body weight, fur, gait). Although, Annex VIII of EU Directive 2010/63 provides examples of each severity category, researchers still lack tools to estimate the degree of severity of a given procedure using scientifically based data.

The absence of indicators showing negative well-being is not the only way to determine the state of the animal; the presence of indicators pointing to positive well-being is also important[5,6,7,8]. For example, animals display luxury behaviors like burrowing and nest building behavior only when all their essential needs are met. If well-being is reduced, luxury behaviors are the first to decline[5,7]. Protocols to be used in assessing

well-being should include indicators pointing to the physical, physiological/biochemical, and psychological states of animals in order to evaluate their well-being in a detailed and comprehensive manner[9].

Within the context of refinement, a protocol was developed to meet these requirements and to assess the effects of a procedure involving general anesthesia on well-being of mice[10]. At the same time, the goal was to minimize any additional stress to enable the easy integration of the protocol into a given experiment. The protocol considers burrowing behavior, home cage behavior such as activity, food intake, and nesting, and trait anxiety-related behavior. In addition, it includes the Mouse Grimace Scale (MGS), and the non-invasive analysis of corticosterone metabolites in feces. The protocol is designed to facilitate the assessment of well-being in a scientific and animal-centered manner and to provide information on well-being that supports the classification of the degree of severity. In addition to score sheets, it can provide useful information for the severity classification of a procedure. As the protocol is easy to carry out and does not require extensive equipment, it can be integrated into an ongoing experiment without influencing the results of a study. It should be noted that the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guideline[11] is to be observed in all studies involving animal experiments, with the goal of improving design, analysis, and reporting.

Protocol

The study was performed according to the guidelines laid down by the German Animal Welfare Act and was approved by the Berlin State Authority ("Landesamt für Gesundheit und Soziales", permit number: G0053/15).

NOTE: The main objective of this protocol was to investigate the effect of repeated anesthesia on glucocorticoid metabolites. A sample size calculation was performed to determine the number of animals to be used: $n \geq 2 \times (s/\mu_1 - \mu_2)^2 \times (z_\alpha + z_\beta)^2$. $\mu_1 - \mu_2$ is the difference between population means at which power and sample size calculations are made ($\alpha = 5\%$, $\beta = 80\%$); $z_\alpha = 1.96$ and $z_\beta = 0.84$ are the quantiles of the standard normal distribution. **Figure 1** illustrates the time line of this protocol. If a parameter of the protocol shows a difference with the control level, the animal should be closely monitored, and the parameter should be measured again after a suitable period. For example, if trait anxiety-related behavior is increased, this behavior should be tested again a week later, in order to help determine the period until full recovery. Time points and periods defined in this protocol can be adapted for use with other procedures. When changing time points, habituation periods should be kept as described in the protocol. In order to reduce factors that might influence the mice's behavior, tests requiring more manipulation should be conducted after tests that do not disturb the normal behavior of mice. **Figure 2** summarizes all tests of the protocol using a summary scoring sheet. **Figure 3** provides simplified scales of the grade of well-being, which give an overview of how to interpret the test results.

1. Habituating mice to handling by experimenter

1. Allow mice to habituate to the animal facility for at least 2 weeks after they have been obtained from another facility or vendor.
2. House mice in groups and maintain them under standard conditions (room temperature 22 ± 2 °C; relative humidity 55 ± 10%) on a light:dark cycle of 12:12 h.
3. Provide all groups with a tunnel and cotton nestlets as standard enrichment, and provide food and water *ad libitum*.
4. Habituate all mice to the tunnel and/or cup handling at least a week prior to testing[12].
   NOTE: Picking up mice by the tail can induce stress or anxiety, which in turn affects well-being and also has an impact on the results of this protocol[12].

2. Preparing the behavioral testing room and apparatuses

NOTE: Provide a separate room for testing, ideally near the room where the animals are kept. Transport the mice in their home cages to the testing room at least 60 min before the procedure is conducted. If possible, conduct all tests of this protocol in the same testing room where the procedure is carried out.

1. Prepare an observation cage to test burrowing behavior[8] and to take the photographs for use in the MGS[13] (**Figure 4**).
   1. Use a glass box with a floor area of approximately 220 mm × 290 mm and a height of 390 mm.
   2. Cover the floor of this box with approximately 0.5 cm of bedding material.
   3. Scatter a handful of used bedding material from the home cage on top of the new bedding material to reduce distress caused by the new environment.
   4. Provide food, the same kind that is normally supplied as diet, and water.
      NOTE:If possible, use water bottles, because mice may fill water bowls with bedding material .

2. Prepare a cage (type III: 420 mm × 260 mm × 150 mm) for the 24-h observation period, for which mice are housed individually (**Figure 5**).
   NOTE: In order to minimize the duration of individual housing, collect data for nest building behavior, home cage activity, food intake, and fecal corticosterone metabolites (FCM) during this period.
   1. Place new bedding material in the cage (approximately 0.5 cm deep) and scatter a handful of used bedding material without feces from the home cage on top of the new material, in order to reduce distress.
   2. Provide a standardized square cotton nestlet of a defined weight, as environmental enrichment only (see **table of materials**)[14].
      NOTE: Commercial nestlets might differ in weight. Therefore, we modified the weight of the nestlet described by Deacon and used 2.0 g instead of 2.7 g[14].
   3. Mount the infrared sensor on the top of the cage, when using an infrared sensor to measure home cage activity (see **table of materials**).
   4. Provide food, the same kind that is normally supplied as diet, and water *ad libitum*.

3. Mouse Grimace Scale

NOTE: Photographs for the MGS are taken in the observation cage at three time points: (i) 2 days prior to the procedure to record baseline MGS levels, (ii) 30 min after the procedure, and (iii) 150 min after the procedure. When well-being is impaired, scores on the MGS increase. If increased MGS scores are still observed after 150 min, take additional photographs at a later stage.

1. Use a high-definition camera for photography.
2. Gently transfer the mouse into the observation cage and allow the mouse to habituate to the new environment for at least 30 min.
3. Continuously take about 30 - 40 photographs for each time point within 1 - 2 min.
4. Sort all photographs by selecting the sharp frontal or lateral photographs and discarding blurry photographs or photographs that show mouse faces from other perspectives than frontal or lateral view.
5. Randomly select one photograph from each time point, (*i.e.* 2 days prior to the procedure, 30 min after the procedure, and 150 min after the procedure) for each mouse.
6. Crop the photographs to display only the head of the mouse so that the body position is not visible[13].
7. Create a spreadsheet file with one sheet for each photograph and add a table including the five facial action units of the MGS to each sheet.
   NOTE: The file contains baseline photographs as well as photographs post procedure.
8. Randomize the order of the sheets.
9. Present the file on a computer screen to three independent persons, who were previously trained to use the MGS developed by Langford *et al.* and have them score the facial action units using a 3-point-scale (0 = not present, 1 = moderately present, 2 = obviously present).
   NOTE: Scoring is based on following parameters[13]: Orbital tightening ("narrowing of the orbital area, with a tightly closed eyelid or an eye squeeze"); nose bulge ("rounded extension of skin visible on the bridge of the nose"); cheek bulge ("convex appearance of the cheek muscle"); ear position ("ears pulled apart and back from their baseline position or featuring vertical ridges that form owing to tips of ears being drawn back"); whisker change ("movement of whiskers from their baseline position either backward, against the face or forward, as if standing on end; whiskers may also clump together").
10. Analyze scores, as follows (adapted from Langford *et al.*[13]).
    1. Average all facial action units for each photograph to generate the MGS score.
       NOTE: If one of the facial action units could not be scored, average the remaining facial action units.
    2. Subtract the mean for the baseline photographs from the mean for the photographs post procedure to obtain a MGS difference score for each mouse.
    3. Test for differences in the MGS difference scores between the persons (nonparametric test for related samples).
       NOTE: If there is a significant difference ($p < 0.05$), determine whether scores of all photographs or only scores of a few photographs differ between the persons. If the latter is true, repeat scoring of these photographs. Otherwise, the persons should repeat the MGS training and then score the photographs again.
    4. Average the MGS difference scores obtained from the different scorers for each mouse, if results of all persons do not significantly differ.
    5. Use a nonparametric statistical test to compare the MGS difference scores averaged between the study groups.

4. Burrowing behavior[8,15,16]

1. Prepare burrows by placing 140 ± 2 g food pellets normally supplied as diet in a standard opaque plastic water bottle (250 mL, 150 mm length, 55 mm diameter, 45 mm diameter of bottle neck)[8].
   NOTE: As mice prefer wide tubes, burrows with a diameter of 68 mm can be used as described by Deacon[16].
2. Place the burrow filled with food pellets in the home cage 5 days prior to the procedure for acclimatization.
   NOTE: The regular food-dispensing unit in the cage should not be emptied but should also remain filled with food pellets, as mice are used to this.
3. Carry out the test twice, 2 days prior to the procedure (baseline); carry out the last 30 min post procedure as well.
   1. Let the mouse habituate for at least 30 min to the observation cage where photographs for the MGS were taken.
   2. Place the plastic water bottle filled with food pellets parallel to the back wall of the observation cage.
   3. Weigh food pellets (g) remaining in the burrow after 2 h.
4. Calculate the weight of food pellets removed from the burrow by mice relative to initial weight (%).

5. 24-h observation period

NOTE: Mice are housed individually, as described in 2.2. (**Figure 5**), for a period of 24 h, in order to measure food intake, home cage activity, nest building behavior, and FCM levels. The 24-h observation takes place twice: (i) 2 days prior to the procedure for baseline levels, (ii) on the day of the procedure.

1. **Food intake**
   1. Weigh the mice at regular intervals (*e.g.* 2 days before anesthesia, immediately before anesthesia, 2 days after anesthesia and weekly after anesthesia), in order to evaluate any changes in body weight (part of the score sheet).
      NOTE: Body weight is required to calculate food intake per gram of body weight. Water intake can also be measured during the 24-h observation period. If food intake is reduced, well-being may be impaired.
   2. Determine the initial weight of standard food diet (grams) provided in the food unit of the cage (approximately 100 g).
   3. Determine the weight of standard food diet at the end of the 24-h observation period.
   4. Scan the cage side beneath the food unit carefully for food spillage and add any extra food pellets found to the weight of food pellets remaining in the food unit.

5. Calculate food intake per unit body weight.

2. **Home cage activity**
NOTE: The following instructions refer to the use of an infrared sensor (see **table of materials**), but home cage activity can also be assessed with alternative programs. Deviation of home cage activity from control levels (*e.g.* hypoactivity, hyperactivity) may be a sign of impaired well-being.
    1. Start the program.
    2. Choose a sample interval of 1 min and an acquisition time of 24 h, meaning that impulses are recorded every minute for 24 h.
    NOTE: If the experimenter enters the room several times after recording started, only use data from periods, when mice were not disturbed (*i.e.* during the dark period).
    3. Sum up 10-min intervals of impulses.
    4. Calculate the area under the time curve (impulses × min).

3. **Nest building behavior**
NOTE: Complex and high nests can serve as an indicator of well-being.
    1. Place a square cotton nestlet (see **Table of Materials**) with a defined weight (*e.g.* 2.0 g) in the middle of the cage.
    2. Score the nest on a 5-point scale (see below) according to Deacon[14] the following morning, approximately 2 h after the light turns on. Weigh any untorn nestlet pieces that are at least 5% of the initial nestlet weight. Score the nests as follows[14]
        1. Assign score of "1" if 90% of the nestlet intact.
        2. Assign score of "2" if it is 50 - 90% intact.
        3. Assign score "3" if 50 - 90% of the nestlet is shredded.
        4. Assign score "4" if more than 90% is shredded but nest is flat, and less than 50% of its circumference is higher than mouse body height when curled up.
        5. Assign score "5" if more than 90% nestlet is shredded and nest is high, and more than 50% of its circumference is higher than body height of a curled up mouse.

4. **Fecal Corticosterone Metabolites**
NOTE: Increases of FCM above the control level reflect acute stress levels over the 24-h postanesthetic period.
    1. Collect all dry fecal pellets from the cage by using forceps at the end of the 24-h observation period and eliminate wet pellets contaminated with urine.
    2. Extract FCM according to Palme *et al.*[17], as follows.
        1. Dry fecal samples at a temperature of 60 - 70 °C.
        2. Homogenize fecal samples using a mortar.
        3. Shake an aliquot of 0.05 g with 1 mL of 80% methanol in a centrifuge tube for 30 min on a multi-vortex.
        4. Centrifuge samples at 2500 x g for 15 min.
        5. Pipette 0.5 mL of supernatant into another centrifuge tube.
        6. Store fecal samples (and extracts) at a minimum of -18 °C.
        7. Analyze FCM using a 5α-pregnane-3b,11b,21-triol-20-one enzyme immunoassay (EIA)[18,19] or another fully validated EIA.
    3. Calculate the percentage change of FCM concentrations relative to the baseline FCM concentrations.

6. Free exploratory paradigm

1. Take the home cage out of the rack and place it on a table surface at the end of the 24-h observation period.
2. Place a gridded cage top (without food or water bottles) in the cage at an angle of 45° to the longer side of the cage.
NOTE: Do not destroy the nest, which serves as a hiding place for the mouse, but place the cage top diagonally above the nest.
3. Monitor or video-record the mice for 10 min from a distance of approximately 1.5 m.
    1. Start the timer.
    2. Note all times when the mouse climbs onto the cage top (with all four paws on the cage top) or leaves the cage top (with one or more paws on the cage floor).
    NOTE: Some mice may climb up the cage top and leave it in order to walk along the edge of the cage. Some mice also rear on the cage top. Treat these cases as if the mice were still on the cage top.

4. Evaluate parameters following Bert *et al.*[20].
    1. Analyze latency to first exploration (in seconds).
    2. Analyze number of explorations.
    3. Analyze total duration (seconds) of exploration.
    NOTE: A high latency to first exploration, a low number of explorations, and a low total duration of exploration can indicate higher trait anxiety levels.

Representative Results

This protocol was originally developed to assess well-being of C57BL/6JRj mice following a single experience of isoflurane anesthesia (one 45-min anesthesia session, n = 13 females) or repeated isoflurane anesthesia (six 45-min anesthesia sessions with 3 - 4 days between the anesthesia sessions, n = 13 females) compared with the well-being of control mice (n = 6 females)[10], which received no anesthesia but were tested according to the same measures. We assessed the impact of a single experience of isoflurane anesthesia and repeated isoflurane anesthesia on the well-being of C57BL/6JRj mice compared with untreated control mice. Here, representative results of female C57BL/6JRj mice, including some data previously published in Hohlbaum *et al.*[10], as well as previously unpublished results, are shown.

Statistical analysis

Explorative data analysis and tests for normality were performed for each parameter. Differences between the study groups (*i.e.* control, single isoflurane anesthesia, repeated isoflurane anesthesia were analyzed using the respective test, as stated in the figure legends. When data met normal distribution assumptions, 1-way ANOVA was performed. Non-normally distributed data were analyzed using the Kruskal-Wallis-Test. Differences were considered significant at $p < 0.05$.

Baseline values

Baseline values, collected before the procedure is carried out, are crucial to determine whether treatment groups differ in the respective parameter. As demonstrated in **Figure 6**, **Figure 7**, **Figure 8**, **Figure 9** and **Figure 10**, baseline levels of the MGS score ($p = 0.762$, Kruskal-Wallis-Test), luxury behaviors like burrowing ($p = 0.896$, Kruskal-Wallis-Test) and nesting ($p = 0.723$, Kruskal-Wallis-Test), food intake ($p = 0.398$, 1-way ANOVA), and home cage activity ($p = 0.208$, Kruskal-Wallis-Test) did not significantly differ between the groups. Moreover, no significant differences in baseline FCM concentrations were found (median, interquartile range in brackets [ng/50 mg]: control:123.01 (82.70 - 193.46); single anesthesia: 118.31 (101.73 - 153.54); repeated anesthesia: 129.55 (92.58 - 139.48)) ($p = 0.904$, Kruskal-Wallis-Test). If differences in the baseline levels occur, delta values can be calculated.

Mouse Grimace Scale

When the mean MGS scores are compared, significant higher scores versus the control were found after the single experience of anesthesia ($p = 0.001$) and after the last repeated anesthesia session ($p = 0.021$) caused at 30 min after last anesthesia (**Figure 6A**). At 150 min after last anesthesia, there were no longer differences between the groups ($p = 0.910$).

To take into account the fact that baseline MGS scores are not equal to 0 in all cases, the MGS difference score was calculated, as described in the protocol. Both a single experience of anesthesia ($p = 0.002$) and repeated anesthesia ($p = 0.008$) increased the MGS difference scores versus the control at 30 min after anesthesia. At 150 min after the last anesthesia, all mice had returned to control levels ($p = 0.617$) (**Figure 6B**)[10].

Burrowing behavior

Repeated anesthesia significantly reduced the percentage of weight of food pellets mice removed from the burrow versus the control ($p = 0.036$, Kruksal-Wallis-Test) (**Figure 7**)[10].

Nest building behavior

There were no significant differences in the nest scores between a single experience of anesthesia, repeated anesthesia, and the control ($p = 0.240$, Kruksal-Wallis-Test) (**Figure 8**)[10].

Food intake

At 1 day after the last anesthesia, mice that had undergone repeated anesthesia showed significantly reduced food intake compared with mice that had experienced single anesthesia ($p = 0.047$, 1-way ANOVA). In contrast, a week later, mice that had received repeated anesthesia consumed significantly more food than controls ($p = 0.012$, 1-way ANOVA) or mice that had received a single anesthesia ($p = 0.001$, 1-way ANOVA) (**Figure 9**)[10].

Home cage activity

At 1 day after the last anesthesia, home cage activity during the dark period, indicated by the area under the activity curve, did not significantly differ among mice that had received a single anesthesia, repeated anesthesia, or control treatment ($p = 0.498$, Kruskal-Wallis-Test) (**Figure 10**)[10].

Free exploratory paradigm

All mice explored the cage top, when the test was carried out. However, 1 day after the last anesthesia, mice that had received repeated anesthesia (median, interquartile range in brackets [s]: 78.00 (55.00-89.00)) explored the cage top significantly later in time than did controls (31.00 (18.25-42.75); p = 0.009, Kruskal-Wallis-Test) and mice that had received a single anesthesia (27.00 (21.00-45.50); p = 0.001, Kruskal-Wallis-Test), as previously published[10]. The parameters total duration of exploration (**Figure 11A**) and number of explorations (**Figure 11B**) are analogous to the latency to first exploration. Repeated anesthesia significantly reduced the number of explorations versus control (p = 0.023, Kruskal-Wallis-Test) and the total duration of exploration (p = 0.032, Kruskal-Wallis-Test) versus a single anesthesia at 1 day after the last anesthesia. At 8 days after the last anesthesia, all parameters, *i.e.* latency to first exploration (control: 27.50 (13.50 - 47.25); single anesthesia: 18.00 (9.50 - 38.50); repeated anesthesia: 20.00 (12.50 - 42.00); p = 0.722, Kruskal-Wallis-Test), number of explorations (p = 0.057), and total duration of exploration (p = 0.579), no longer differed between the study groups (**Figure 11**).

Fecal corticosterone metabolites

In order to take into account the baseline values obtained, the percentage change relative to baseline was calculated and no significant differences between the groups were found (p = 0.119, Kruskal-Wallis-Test) (**Figure 12**)[10].

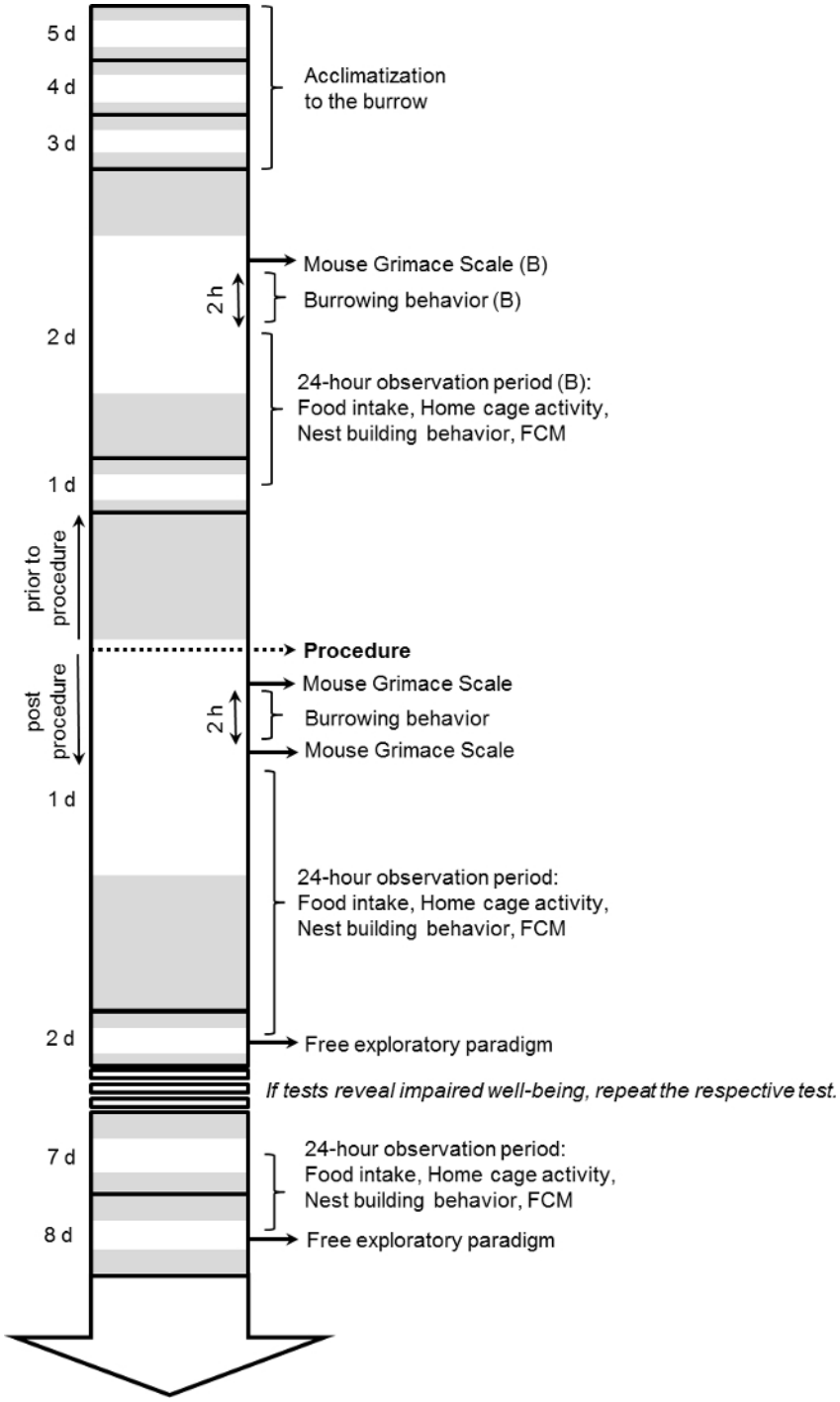

Figure 1: Time line of protocol. Gray and white colored fields symbolize the dark and light periods of a day, respectively. Depending on the procedure, this protocol may be adapted. "Habituation to the burrow" means that mice have to be acclimatized to using the plastic water bottle as a burrow in their home cage, before the burrowing test may be conducted. B, baseline value; FCM, fecal corticosterone metabolites. This figure has been modified from Hohlbaum *et al.*[10]. Please click here to view a larger version of this figure.

Summary Scoring Sheet

Mouse ID: _____
Experimentator(s):_____

Mouse Grimace Scale (MGS)	Mean MGS score [0–2]			Mean MGS difference score		
	Baseline (date/time)	**30 min post anesthesia** (date/time)	**150 min post anesthesia** (date/time)	**30 min post anesthesia** (date/time)	**150 min post anesthesia** (date/time)	**Average mean MGS difference scores of all scorers.**
Scorer 1						
Scorer 2						
Scorer 3						

Burrowing Behavior	**Baseline** (date/time)			**Post-procedure** (date/time)		
	Initial weight of pellets [g]	Weight of pellets remaining in the burrow [g]	% of pellets removed from burrow relative to initial weight [%]	Initial weight of pellets [g]	Weight of pellets remaining in the burrow [g]	% of pellets removed from burrow relative to initial weight [%]
Weight of pellets						

24-hour observation period	**1 day post-procedure** (date/time)	**8 days post-procedure** (date/time)
Nest building behavior [score 1–5]		
Food intake/g body weight [g]		
Home cage activity, area under the curve [impulses × min]		
Fecal corticosterone metabolites [ng/50mg]		

Free exploratory paradigm	**1 day post-procedure** (date/time)	**8 days post-procedure** (date/time)
Latency to first exploration [s]		
Number of explorations [n]		
Total duration of exploration [s]		

Figure 2: Summary scoring sheet. This sheet includes all tests and can be filled out for each individual mouse. Date and time should be added when the test is conducted. Please click here to view a larger version of this figure.

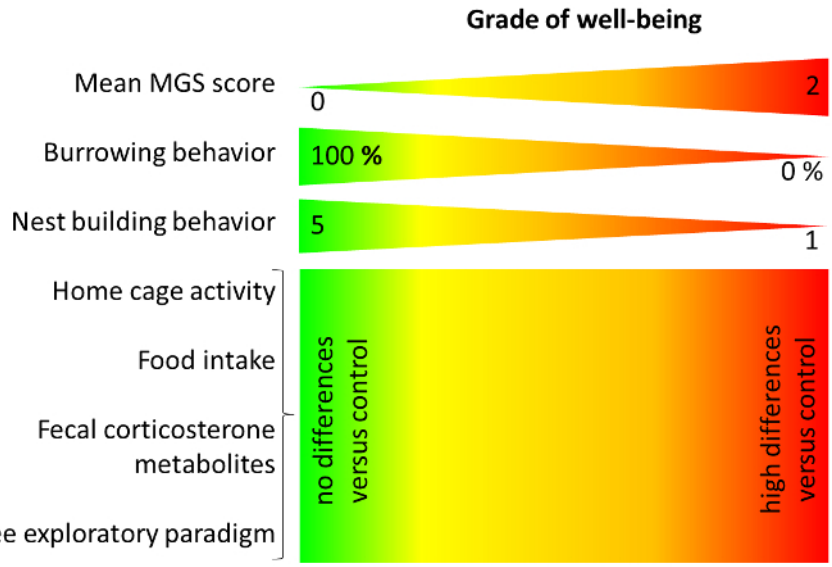

Figure 3: Scales of the grade of well-being. The scales range from "well-being unimpaired" (green) to "well-being impaired" (red) and express the meaning of the test results in a simplified way. At this stage of knowledge on indicators for well-being, we cannot make a definite statement on the grade of well-being for each test, only a general statement. MGS, Mouse Grimace Scale. Please click here to view a larger version of this figure.

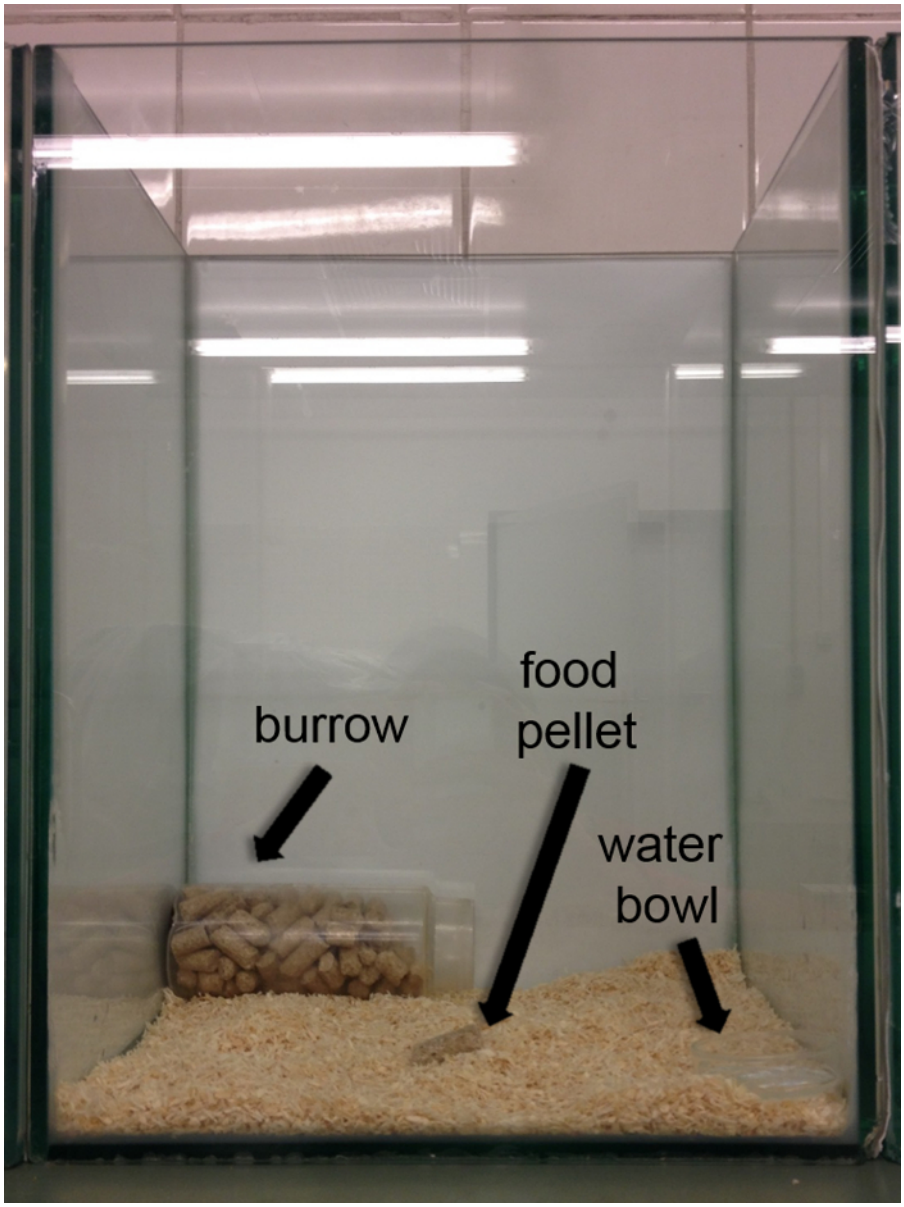

Figure 4: Observation cage. This cage is used for the test of burrowing behavior and to take photographs to be analyzed according to the Mouse Grimace Scale (MGS). The front of the cage is clear and, depending on the fur colors of the mice, the other three walls should be colored either black or white to contrast with the mice. The floor of the glass box is covered with approximately 0.5 cm of bedding material including used bedding material from the home cage. Standard food and water are provided. If possible, water bottles rather than bowls should be used, because mice may fill bowls with bedding material. After a habituation period of at least 30 min, the burrow is added. Please click here to view a larger version of this figure.

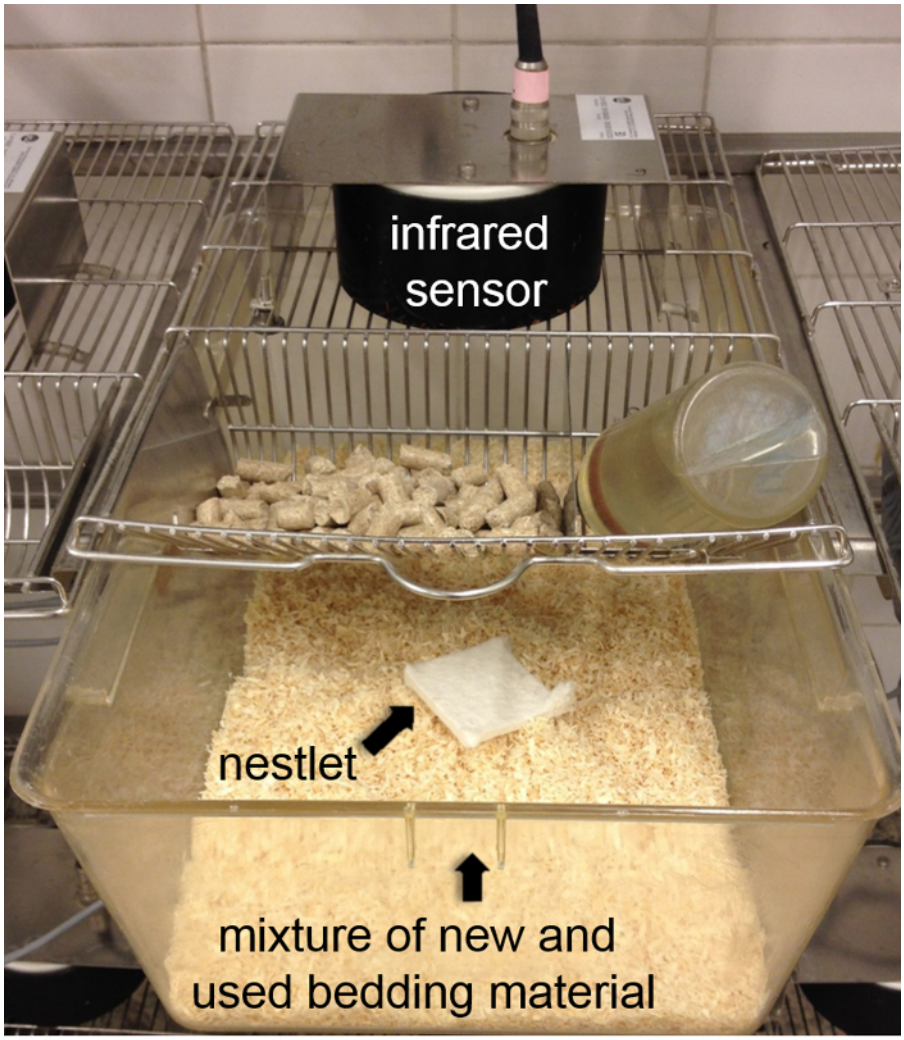

Figure 5: 24-h observation period. To assess nest building behavior, home cage activity, and food intake and to collect fecal samples from which to measure fecal corticosterone metabolites (FCM), mice are housed individually for 24 h (cage type III: 420 mm × 260 mm × 150 mm; bedding material approximately 0.5 cm deep with used bedding material from the home cage scattered on top, a cotton nestlet, tap water and standard food diet *ad libitum*). An infrared sensor is mounted on the top of the cage. Please click here to view a larger version of this figure.

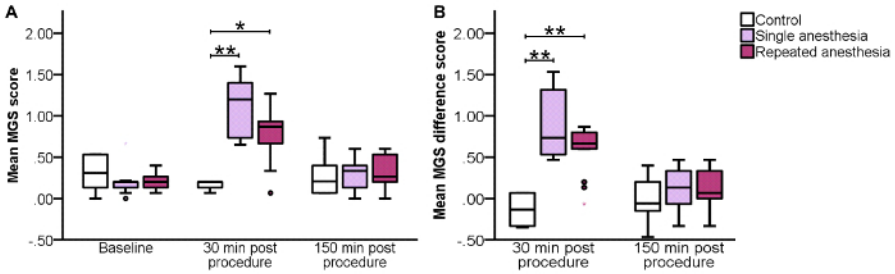

Figure 6: Mouse Grimace Scale (MGS). The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile. The whiskers represent values which are no greater than 1.5 × IQR. Dots are outliers with values between 1.5 - 3.0 × IQR. Colored asterisks are outliers with values greater than 3.0 × IQR. (**A**) Mean MGS score. (**B**) Mean MGS difference score. p values were calculated using Kruskal-Wallis-Test: *p < 0.05; **p < 0.01. Due to a technical malfunction (camera, scale), four mice in the single anesthesia group had to be excluded from the statistics. This figure has been modified from Hohlbaum *et al.*[10]. Please click here to view a larger version of this figure.

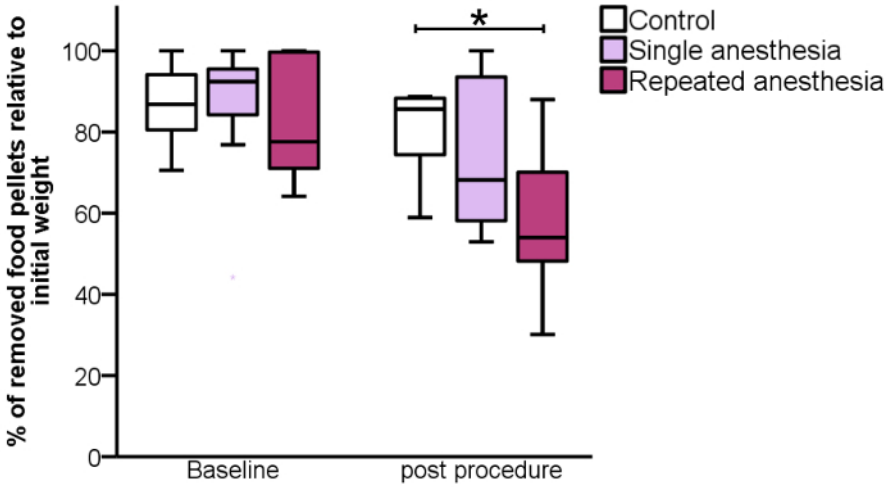

Figure 7: Burrowing behavior. The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile. The whiskers represent values which are no greater than 1.5 × IQR. Colored asterisks are outliers with values greater than 3.0 × IQR. p values were calculated using Kruskal-Wallis-Test: *p < 0.05. Due to a technical malfunction (camera, scale), four mice in the single anesthesia group had to be excluded from the statistical analysis. This figure has been modified from Hohlbaum *et al.*[10]. Please click here to view a larger version of this figure.

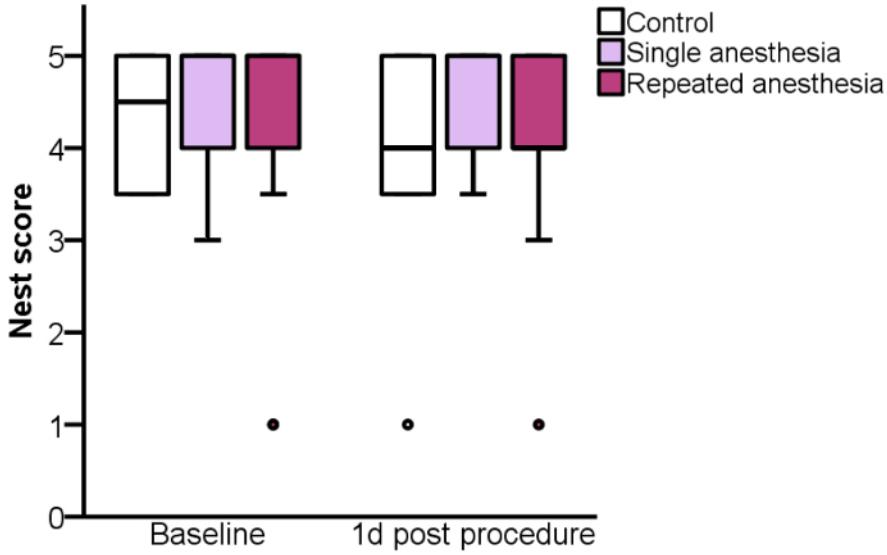

Figure 8: Nest building behavior. The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile. The whiskers represent values which are no greater than 1.5 × IQR. Dots are outliers with values between 1.5 - 3.0 × IQR. p values were calculated using Kruskal-Wallis-Test; d, day. Please click here to view a larger version of this figure.

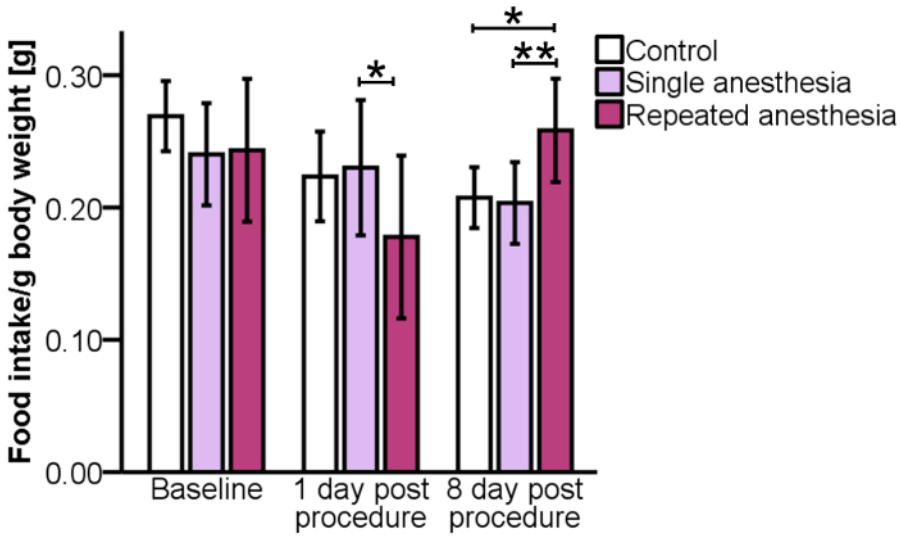

Figure 9: Food intake. Data are mean ± standard deviation. p values were calculated using 1-way ANOVA (post-hoc Tukey-HSD): *p < 0.05; **p < 0.01. Please click here to view a larger version of this figure.

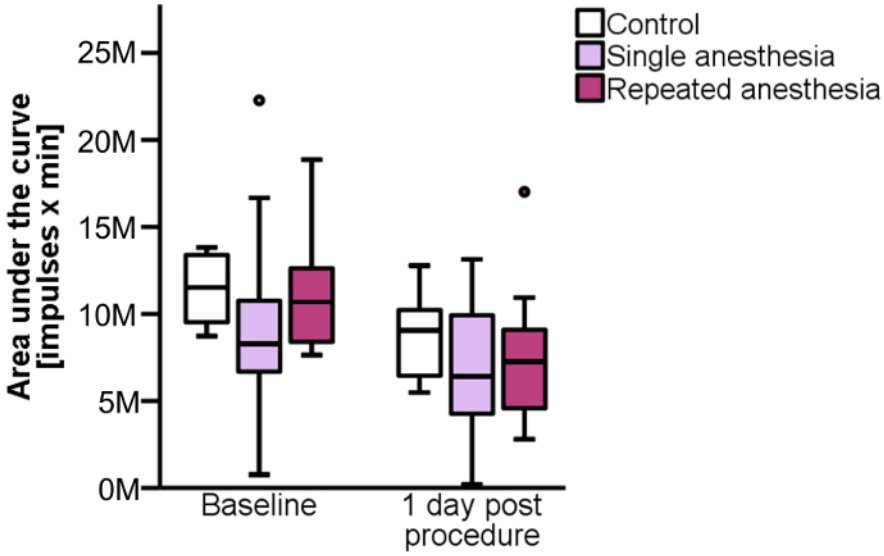

Figure 10: Home cage activity. The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile. The whiskers represent values which are no greater than 1.5 × IQR. Dots are outliers with values between 1.5 - 3.0 × IQR. p values were calculated using Kruskal-Wallis-Test. Please click here to view a larger version of this figure.

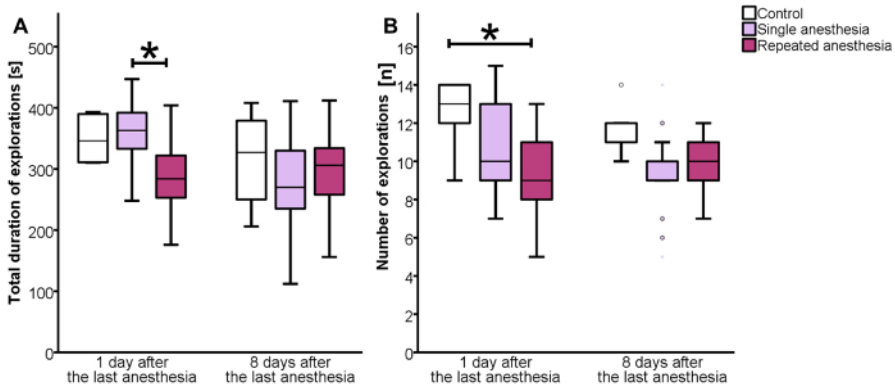

Figure 11: Free exploratory paradigm. The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile. The whiskers represent values which are no greater than 1.5 × IQR.Dots are outliers with values between 1.5 - 3.0 × IQR. Colored asterisks are outliers with values greater than 3.0 × IQR. (A) Total duration of exploration. (B) Number of explorations. p values were calculated using Kruskal-Wallis-Test: * $p < 0.05$. This figure has been modified from Hohlbaum *et al.*[10]. Please click here to view a larger version of this figure.

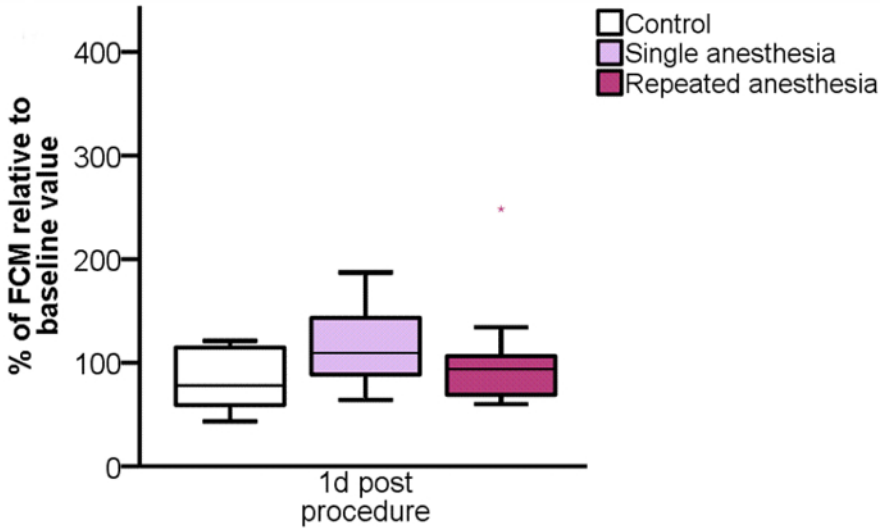

Figure 12: Fecal corticosterone metabolites (FCM). The box represents the interquartile range (IQR), box edges are the 25th and 75th quartile. The whiskers represent values which are no greater than 1.5 × IQR. Colored asterisks are outliers with values greater than 3.0 × IQR. FCM levels were measured 2 days prior to and 1 day post procedure. The percentage change [%] of FCM concentrations relative to respective baseline value were calculated. Data were analyzed using Kruskal-Wallis-Test. Please click here to view a larger version of this figure.

Discussion

The protocol was originally developed to assess well-being of C57BL/6JRj mice that received either a single anesthesia or repeated isoflurane anesthesia. The results confirm that tests of luxury behaviors, as well as other measures (*e.g.* the free exploratory paradigm, the MGS, burrowing food intake) were sensitive methods for assessing well-being. Repeated isoflurane anesthesia caused short-term effects on trait anxiety-related behavior, the MGS, and burrowing behavior. Moreover, repeated isoflurane anesthesia appeared to influence food intake[10].

The free exploratory paradigm indicated reduced exploratory behavior and, consequently, higher trait anxiety levels in mice that were repeatedly anesthetized compared with mice anesthetized only once, and the controls. However, mice of all groups explored the gridded cage top[10,20]. When investigating anxiety-related behavior, it is important to distinguish between trait and state anxiety[21,22]. Placing the mouse in an unfamiliar environment induces state anxiety. In contrast, the free exploratory paradigm allows the animals to stay in their home cages and, thus, investigates trait anxiety.

The reduction in exploratory behavior among mice after repeated anesthesia cannot be explained by decreased motor activity. Home cage activity during the dark period did not significantly differ between the study groups. This indicates that mice had already physically recovered from the anesthesia when the free exploratory paradigm test was performed.

The MGS was actually developed to assess pain, but there is evidence that the facial expressions are also altered by stress and positive emotions[13,23]. The photographs of the mouse faces indicated that both single and repeated isoflurane anesthesia increased scores on the MGS over the short-term in the immediate post-anesthetic period. As the MGS difference score remained below 1, the level of stress appeared to be mild[10]. The results agree with recent observations of Miller *et al.*[24,25] and showed that mice returned to control level 150 min post anesthesia[10].

Luxury behaviors like burrowing and nest building behavior are species-specific and can serve as indicators for well-being[7] and good general health in mice[26]. Mice only show luxury behaviors when all their essential needs are met. If well-being is reduced, luxury behaviors are the first to be impaired[5,7]. Previous studies have shown that both burrowing and nest building can be impaired by pain and distress[6,8], but there is also evidence that hippocampal lesions can affect these two behaviors[15,27,28,29,30,31].

Luxury behaviors were investigated in the early post-anesthetic period (burrowing behavior) and by the morning of the following day (nest building behavior). The test for burrowing behavior developed by Deacon *et al.* and adopted by Jirkof *et al.* was modified with regard to acclimatization (group housing instead of individual housing) and duration of behavioral measurement (only 2 h instead of 24 h)[8,15,16].

It is notable that repeated anesthesia reduced burrowing behavior, suggesting that there was an impairment of well-being immediately post anesthesia[10], which was also reported by Jirkof *et al.*[8]. But in the morning of the following day, when mice had had more time to recover from anesthesia, the high and complex nests they had built indicated that they were experiencing well-being[10]. These findings differ from previous reports that scored nests at an earlier time[32]. Therefore, it may be useful to adapt the protocol and score nests earlier in time. However, the circadian rhythm of nest building behavior must still be considered, as mice tend to prepare their nests at the end of the dark phase[32].

Food intake was marginally decreased 1 day after repeated anesthesia, but it was increased one week later. As mice did not lose body weight (see Hohlbaum *et al.*[10]) some kind of compensation mechanism may be occurring, and impairment of well-being with regard to food intake should be classified as mild[10]. Food intake provides insight into postoperative distress, well-being and appetite in mice, which can be impaired by postoperative nausea[26] and postoperative stress[27].

FCM reliably indicate stress in mice[33,34,35]. The peak FCM concentrations typically occur 8-10 h after a stressor but depend on the intestinal transit time[18]. Therefore, it is crucial to also monitor home cage activity, as included in this protocol. Due to the circadian rhythm effect on excreted FCM[18], it is advisable to collect fecal samples over a period of 24 h. FCM concentrations reflected acute stress over the 24-h post-anesthetic period. As FCM levels vary from individual to individual, the percentage change of FCM levels relative to baseline was calculated. The FCM results of the present investigation indicated that neither one experience of anesthesia nor repeated anesthesia significantly increased hypothalamic-pituitary-adrenal (HPA) axis activity[10].

All in all, the findings revealed that repeated isoflurane anesthesia caused short-term mild distress and impaired well-being in the early postanesthetic period slightly more than did a single experience of isoflurane anesthesia[10].

A protocol to assess well-being should not impose additional distress on the animals. A limitation of the present protocol is that it includes individual housing during the 24-h observation period, which is known to increase plasma corticosterone levels[36]. However, individual housing is necessary to collect valid data for individuals with respect to home cage activity, food intake, nest building behavior, and FCM levels. The duration of individual housing was minimized by investigating those four parameters at the same time (*i.e.* during the 24-h observation period). To prevent the results from being biased by individual housing, control mice underwent the same tests and their results were taken into account. If methods are available to measure those parameters in a group-housing environment, they should be used (*e.g.*an automated home-cage analysis system for home cage activity[37] and for food intake). However, automated home cage analysis systems require additional equipment, which may not be available. In studies focusing on groups of animals instead of individuals, home cage activity, food intake, nest building behavior, and FCM levels may also be evaluated for a group of mice. The same applies to the free exploratory paradigm, which only requires the mice to be visibly marked so that they may be distinguished. If well-being is reduced in a group, all mice in the group should be checked carefully (using the score sheet and clinical examination) to identify the mouse or mice concerned. Further observations at the group level are needed to determine the sensitivity of group values.

With regard to refinement, this protocol can be adapted to and integrated into ongoing studies to assess the impact of a particular procedure on the well-being of mice. Depending on the particular procedure and study design, the most suitable tests from this protocol should be chosen. Here, the burrowing and nest building test, MGS, free exploratory paradigm and measurement of food intake appeared to be particularly beneficial.As behavior of mice follows circadian rhythms[38], it is advisable to perform behavioral tests in mice of all groups at a defined time point. Mice are more active in the morning than in the afternoon[38]. However, if the study design does not allow for testing in the morning, the test may be done at another time. It is also important to ensure that each single test is conducted at the same time points for all groups. Otherwise, the impact of the circadian rhythm of the parameters can result in intra- or intergroup differences. Moreover, control mice should be included in the study, so that the results from treated mice and control mice may be compared. Control mice must be tested at the same time point as treated mice. If the protocol shows that well-being is impaired, the procedure should be refined, and the protocol repeated. This approach can show whether the efforts to refine the procedure were effective.

The present protocol can serve as a general aid to estimate the degree of severity caused by a procedure. Therefore, it helps to classify the severity of a procedure on a scientific- and animal-centered level. However, the protocol does not provide a scale for classifying a procedure as mild, moderate, or severe. To classify the severity of a procedure, it is necessary to refer to the examples in Annex VIII of EU Directive 2010/63.

In conclusion, the present protocol supports a systematic assessment of well-being in mice in a scientific, animal-centered manner following procedures during which general anesthesia is used.

Disclosures

The authors have nothing to disclose.

Acknowledgements

Thanks to Sabine Jacobs for assisting with the sample collection, Edith Klobetz-Rassam for analysis of FCM, PD Dr. med. vet. habil. Roswitha Merle for assisting with statistical analysis, and Wiebke Gentner for proofreading the manuscript. The study is part of the Berlin-Brandenburg

research platform BB3R (www.bb3r.de) and was funded by the German Federal Ministry of Education and Research (grant number: 031A262A) (www.bmbf.de/en/index.html).

References

1. 2010 EU. Directive 2010/63/EU. *Official Journal of the European Union.* **L276/33-L276/29** (2010).
2. Russell, W., Burch, R. The principles of humane experimental technique. London: *Methuen.* (1959).
3. Morton, D. B., Griffiths, P. H. Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *Vet Rec.* **116** (16), 431-436 (1985).
4. Bugnon, P., Heimann, M., Thallmair, M. What the literature tells us about score sheet design. *Lab Anim.* **50** (6), 414-417 (2016).
5. Boissy, A. *et al.* Assessment of positive emotions in animals to improve their welfare. *Physiol Behav.* **92** (3), 375-397 (2007).
6. Arras, M., Rettich, A., Cinelli, P., Kasermann, H. P., Burki, K. Assessment of post-laparotomy pain in laboratory mice by telemetric recording of heart rate and heart rate variability. *BMC Vet Res.* **3**, 16 (2007).
7. Jirkof, P. Burrowing and nest building behavior as indicators of well-being in mice. *J Neurosci Methods.* **234,** 139-146, (2014).
8. Jirkof, P. *et al.* Burrowing behavior as an indicator of post-laparotomy pain in mice. *Front Behav Neurosci.* **4**, 165, (2010).
9. Hawkins, P. *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.* **45** (1), 1-13, (2011).
10. Hohlbaum, K., Bert, B., Dietze, S., Palme, R., Fink, H., Thöne-Reineke, C. Severity classification of repeated isoflurane anesthesia in C57BL/6JRj mice-Assessing the degree of distress. *PLoS ONE.* **12** (6), e0179588, (2017).
11. Kilkenny, C., Browne, W. J., Cuthill, I. C., Emerson, M., Altman, D. G. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8** (6), e1000412, (2010).
12. Hurst, J. L., West, R. S. Taming anxiety in laboratory mice. *Nat Methods.* **7** (10), 825-826, (2010).
13. Langford, D. J. *et al.* Coding of facial expressions of pain in the laboratory mouse. *Nat Methods.* **7** (6), 447-449, (2010).
14. Deacon, R. M. Assessing nest building in mice. *Nat Protoc.* **1** (3), 1117-1119, (2006).
15. Deacon, R. M., Raley, J. M., Perry, V. H., Rawlins, J. N. Burrowing into prion disease. *Neuroreport.* **12** (9), 2053-2057 (2001).
16. Deacon, R. M. Burrowing in rodents: a sensitive method for detecting behavioral dysfunction. *Nat Protoc.* **1** (1), 118-121, (2006).
17. Palme, R., Touma, C., Arias, N., Dominchin, M. F., Lepschy, M. Steroid extraction: get the best out of faecal samples. *Wien Tierarz Monats.* **100** (9-10), 238-246 (2013).
18. Touma, C., Palme, R., Sachser, N. Analyzing corticosterone metabolites in fecal samples of mice: a noninvasive technique to monitor stress hormones. *Horm Behav.* **45** (1), 10-22 (2004).
19. Touma, C., Sachser, N., Mostl, E., Palme, R. Effects of sex and time of day on metabolism and excretion of corticosterone in urine and feces of mice. *Gen Comp Endocrinol.* **130** (3), 267-278 (2003).
20. Bert, B., Schmidt, N., Voigt, J. P., Fink, H., Rex, A. Evaluation of cage leaving behaviour in rats as a free choice paradigm. *J Pharmacol Toxicol Methods.* **68** (2), 240-249, (2013).
21. Lister, R. G. Ethologically-based animal models of anxiety disorders. *Pharmacol Ther.* **46** (3), 321-340 (1990).
22. Belzung, C., Berton, F. Further pharmacological validation of the BALB/c neophobia in the free exploratory paradigm as an animal model of trait anxiety. *Behav Pharmacol.* **8** (6-7), 541-548 (1997).
23. Finlayson, K., Lampe, J. F., Hintze, S., Wurbel, H., Melotti, L. Facial indicators of positive emotions in rats. *PLoS ONE.* **11** (11), e0166446, (2016).
24. Miller, A., Kitson, G., Skalkoyannis, B., Leach, M. The effect of isoflurane anaesthesia and buprenorphine on the mouse grimace scale and behaviour in CBA and DBA/2 mice. *Appl Anim Behav Sci.* **172** 58-62, (2015).
25. Miller, A. L., Golledge, H. D., Leach, M. C. The influence of isoflurane anaesthesia on the rat grimace scale. *PLoS ONE.* **11** (11), e0166652, (2016).
26. Deacon, R. Assessing burrowing, nest construction, and hoarding in mice. *J Vis Exp.* (59), e2607, (2012).
27. Felton, L. M., Cunningham, C., Rankine, E. L., Waters, S., Boche, D., Perry, V. H. MCP-1 and murine prion disease: separation of early behavioural dysfunction from overt clinical disease. *Neurobiol Dis.* **20** (2), 283-295, (2005).
28. Deacon, R. M., Croucher, A., Rawlins, J. N. Hippocampal cytotoxic lesion effects on species-typical behaviours in mice. *Behav Brain Res.* **132** (2), 203-213 (2002).
29. Filali, M., Lalonde, R., & Rivest, S. Subchronic memantine administration on spatial learning, exploratory activity, and nest-building in an APP/PS1 mouse model of Alzheimer's disease. *Neuropharmacology.* **60** (6), 930-936, (2011).
30. Guenther, K., Deacon, R. M., Perry, V. H., Rawlins, J. N. Early behavioural changes in scrapie-affected mice and the influence of dapsone. *Eur J Neurosci.* **14** (2), 401-409 (2001).
31. Deacon, R. M., Reisel, D., Perry, V. H., Nicholas, J., Rawlins, P. Hippocampal scrapie infection impairs operant DRL performance in mice. *Behav Brain Res.* **157** (1), 99-105, (2005).
32. Jirkof, P. *et al.* Assessment of postsurgical distress and pain in laboratory mice by nest complexity scoring. *Lab Anim.* **47** (3), 153-161, (2013).
33. Atanasov, N. A., Sargent, J. L., Parmigiani, J. P., Palme, R., Diggs, H. E. Characterization of train-induced vibration and its effect on fecal corticosterone metabolites in mice. *J Am Assoc Lab Anim Sci.* **54** (6), 737-744 (2015).
34. Voigt, C. C. *et al.* Hormonal stress response of laboratory mice to conventional and minimally invasive bleeding techniques. *Anim Welf.* **22** (4), 449-455, (2013).
35. Walker, M. K. *et al.* A less stressful alternative to oral gavage for pharmacological and toxicological studies in mice. *Toxicol Appl Pharmacol.* **260** (1), 65-69, (2012).
36. Miyashita, T. *et al.* Social stress increases biopyrrins, oxidative metabolites of bilirubin, in mouse urine. *Biochem Biophys Res Commun.* **349** (2), 775-780, (2006).
37. Bains, R. S. *et al.* Analysis of individual mouse activity in group housed animals of different inbred strains using a novel automated home cage analysis system. *Front Behav Neurosci.* **10** (106), (2016).
38. Saibaba, P., Sales, G. D., Stodulski, G., Hau, J. Behaviour of rats in their home cages: daytime variations and effects of routine husbandry procedures analysed by time sampling techniques. *Lab Anim.* **30** (1), 13-21, (1996).

