ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Grading animal distress and side effects of therapies

Simone Kumstel,[1] [iD] Edgar H. U. Wendt,[1] Johanna Eichberg,[1] Steven R. Talbot,[2] Christine Häger,[2] Xianbin Zhang,[1] Ahmed Abdelrahman,[1] Maria Schönrogge,[1] Rupert Palme,[3] André Bleich,[2] Brigitte Vollmar,[1] and Dietmar Zechner[1]

[1]Rudolf-Zenker Institute of Experimental Surgery, University Medical Center, Rostock, Germany. [2]Institute for Laboratory Animal Science, Hannover Medical School, Hanover, Germany. [3]Unit of Physiology, Pathophysiology and Experimental Endocrinology, Department of Biomedical Sciences, University of Veterinary Medicine, Vienna, Austria

Address for correspondence: Simone Kumstel, Rudolf-Zenker Institute of Experimental Surgery, University Medical Center, Schillingallee 69a, 18057 Rostock, Germany. simone.kumstel@uni-rostock.de

In order to combine high-quality research with minimal harm to animals, a prospective severity assessment for animal experiments is legally required in many countries. In addition, an assessment of the evidence-based severity level might allow realistic harm–benefit analysis and the appraisal of refinement methods. However, only a few examples describe the distress of animals by simple, cost-efficient, and noninvasive methods. We, therefore, evaluated the severity of an orthotopic mouse model for pancreatic cancer using C57BL/6J mice when pursuing two different chemotherapies. We assessed fecal corticosterone metabolites, body weight, distress score, and burrowing, as well as nesting activity. Moreover, we established a multifactorial model using multivariate logistic regression to describe animal distress. This multifactorial analysis revealed that metformin + galloflavin treatment caused higher distress than metformin + α-cyano-4-hydroxycinnamate therapy. Similar results were obtained by using the best cutoff calculated by Youden's *J* index when using only single parameters, such as burrowing activity or fecal corticosterone metabolite concentration. Thus, the present study revealed that single readout parameters, as well as multivariate analysis, can help to assess the severity of animal experiments and detect side effects of therapies.

Keywords: distress analysis; severity assessment; pancreatic cancer; chemotherapy

Introduction

Animal welfare must be considered as an important aspect of biomedical research, because of its implications on moral issues and the quality of science.[1] In order to combine a high quality of research with minimal harm to animals, multiple regulations have been implemented in many countries during the last decades. When planning *in vivo* experiments, scientists should consider the 3R concept (replacement, reduction, and refinement) and conduct an extensive harm–benefit analysis based on the severity of an animal model.[2] Some prospective severity classifications of specific interventions are defined in the European Union (EU) Directive 2010/63/EU in annex VIII. This annex describes certain procedures, which can be classified as "nonrecovery," "mild," "moderate," or "severe." In addition, some

EU guidance documents provide suggestions on how to meet legal requirements[3] of this directive. However, the annex VIII only provides a few examples for severity classification, and mostly lists procedural steps, rather than full procedures or the distress of entire disease models.[4] Moreover, only a few publications give some additional examples[5] for assessing the severity of animal experiments. Therefore, it is necessary to implement methods, which allow us to define the level of distress caused by widely used interventions and specific animal experiments.[5,6] An evidence-based evaluation will allow a realistic harm–benefit analysis and the use of appropriate refinement methods, which might lead to a sustainable reduction of animal suffering.[6]

The hypothetical ideal level of welfare might be accomplished when the nutritional, environmental, health, behavioral, and mental needs of laboratory

animals are met.[7] However, different stressors or interventions, such as handling, surgery, or injections, might provoke short time stress responses or even distress when an animal is unable to cope with a stressor within a distinct period of time.[8] Distress can be caused by pain, but it can also be caused, for example, by anxiety. Therefore, it has been suggested that minimizing distress will be more important for improving the welfare of animals than only focusing on pain reduction. In this context, distress refers to any mental unpleasant level of stress without differentiating between distinct causes. To reliably detect any deviation of the ideal state of welfare, which might lead to distress, the British Working Group on Refinement suggested to analyze parameters, which describe the physical, physiological, and psychological states of the animals.[6]

The physical state of animals can be described by parameters, such as body weight (BW) and posture, as well as the physiological state, such as heart rate or stress hormones.[6] In addition, distress can influence their natural behavior. The latter can be quantitatively assessed by the analysis of burrowing and nesting activity,[9–11] whereas the physiological stress response of animals is often characterized by analyzing the corticosterone concentration in the blood or its metabolites in feces (fecal corticosterone metabolites (FCMs)).[12–15]

The aim of our current study was to evaluate a method for multifactorial distress analysis. A variety of physical, behavioral, and hormonal parameters for animal welfare were assessed in male C57BL/6J mice bearing pancreatic cancer (PC) and undergoing two distinct chemotherapies. The application of the most powerful parameters in a 2D scatter plot followed by multivariate logistic regression was used as a multifactorial model. This method was further compared with the performance of single parameters in order to evaluate and score animal distress.

Materials and methods

Ethical statement

All animal experiments were approved by the German local authority: Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei Mecklenburg-Vorpommern (7221.3-1-019/15), in accordance with the German animal protection law and the EU Guideline 2010/63/EU.[4] The experiments are reported according to the Arrive

Guidelines.[16] Breeding pairs of C57BL/6J mice were originally purchased from Charles River Laboratories and further bred in our facility of the University Medical Center in Rostock under specific pathogen-free conditions. During the experiment, the mice were kept single-housed in type III cages (Zoonlab GmbH, Castrop-Rauxel, Germany) at 12-h dark:light cycle, the temperature of 21 ± 2 °C, and relative humidity of 60 ± 20% with food (pellets, 10 mm, ssniff-Spezialdiäten GmbH, Soest, Germany) and tap water ad libitum; and enrichment was provided by nesting material (shredded tissue paper, Verbandmittel GmbH, Frankenberg, Deutschland), paper roll (75 × 38 mm, H 0528–151, ssniff-Spezialdiäten GmbH), and a wooden stick (40 × 16 × 10 mm, Abedd, Vienna, Austria).

Syngeneic orthotopic PC model

For orthotopic injection of PC cells, 26 male C57BL/6J mice aged 16.7 (14–18.4) weeks (median/interquartile range) and with an average BW of 27 (25.3–27.7) g (median/interquartile range) were anesthetized in the laboratory with 1.2–2.0% isoflurane and a single subcutaneous injection of 5 mg/kg carprofen (Rimadyl®, Pfizer GmbH, Berlin, Germany) was applied 5 min before surgery as analgesia. Isoflurane was chosen because it allowed a fast recovery from anesthesia. The eyes were kept wet by eye ointment. The abdomen of the mice was shaven and disinfected, and the abdominal cavity was opened by laparotomy, and 5 µL of the cell suspension (murine cell line 6606PDA, 2.5 × 10^5/5 µL cells in matrigel) was injected slowly with a 25-µL syringe (Hamilton Syringe, Reno, NV) into the pancreas. The pancreas was placed back into the cavity and the peritoneum was closed by a coated 5-0 Vicryl® suture (Johnson & Johnson Medical GmbH, New Brunswick, NJ). The skin was sewn with a 5-0 Prolene® suture (Johnson & Johnson Medical GmbH) and the mice were placed in front of a heating lamp. The surgery lasted for 15–20 min for each mouse. The chemotherapeutic treatment of the mice started 4 days after tumor cell injection until day 37. Twenty-six mice were allocated in a nonrandom manner, matching the performance of the preexperimental behavior tests (burrowing and nesting) between the sham and treatment groups. Seven mice aged 18.3 (18.3–21.3) weeks (median/interquartile range) and with a BW of 27 (26.3–27.7) g (median/interquartile

range) received the combinatorial chemotherapy of metformin (Met; 125 mg/kg in phosphate-buffered saline (PBS), daily; Merck, Darmstadt, Germany) and CHC (15 mg/kg in 50% dimethyl sulfoxide (DMSO), daily; Tocris Bioscience, Bristol, UK). The corresponding sham treatment (PBS, 50% DMSO) was performed on an additional seven mice, which were 18.3 (18.3–21.4) weeks old (median/interquartile range) and weighted 28 (27.5–28.5) g (median/interquartile range). The second therapeutic intervention Met (125 mg/kg in PBS; daily) in combination with galloflavin (Gallo) (20 mg/kg in 100% DMSO, three times a week; Tocris Bioscience) was performed on seven mice of an average age of 13.6 (13.4–14.0) (median/interquartile range) and the initial BW of 25 (23.1–26.3) g (median/interquartile range), and five mice of an exact age of 14 weeks and the average BW of 26 (24.4–26.4) g (median/interquartile range) were treated with the respective vehicle (PBS, 100% DMSO). The unequal group size was caused by the termination of the project because the original goal to see an effect on tumor weight was not met. An overview of the number of animals used for each figure is illustrated in Figure S1 (online only). One scientist administered the therapy and another scientist evaluated the distress in a nonblinded fashion. Met and PBS were injected in the morning between 9:00 and 11:00 a.m., while Gallo, CHC, or the corresponding vehicle (100% DMSO, 50%) were injected between 2:30 and 3:00 p.m. In addition to the single intraoperative injection of carprofen (5 mg/kg), 1250 mg/L metamizol (Ratiopharm, Ulm, Germany) was provided daily in the drinking water during the whole experimental period. This multimodal analgesia regime proved to be useful for this animal model in previous studies.[17,18] The benefit of metamizol is that it is self-applicable via drinking water, and it is known to be an effective analgesic for gastrointestinal diseases in mice.[19]
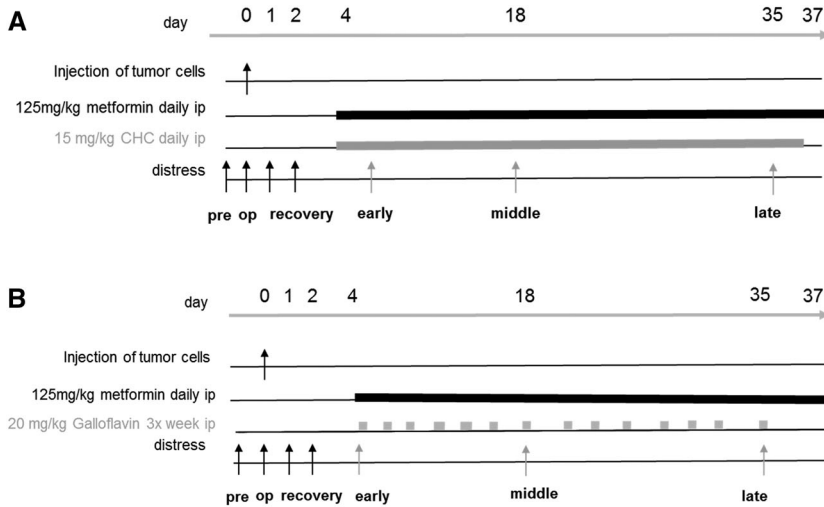
Assessment of distress parameters
To analyze animal distress upon laparotomy and tumor cell injection, all distress parameters were assessed before (pre) and directly after operation (op) until recovery day 2. During chemotherapy, all distress parameters were quantified during the early (days 4–8), middle (days 18–19), and late (days 34–35) phases of therapy, to exemplarily analyze the course of distress during the treatments

(Fig. 1). This time schedule proved to be useful for various projects, for example, for comparing distress between different animal models.[18]

The BW was determined 24 h after surgery or last injection of therapeutic intervention in order to allow enough time for a BW adjustment. The percentage of BW change for all days was always calculated referring to the BW measured before surgery. The distress score was always assessed according to a scoresheet, which was previously published by our working group[20,21] and based on other scoresheets.[22,23] The mice were, therefore, observed in their home cage for a few minutes and a score was assessed when one or more defined criteria were diagnosed concerning the general condition, spontaneous behavior, flight behavior, or process-specific criteria. The distress score was assessed 30 min after the last injection of compounds (Gallo, CHC, or the appropriate vehicle) around 3:00–3:30 p.m.

The burrowing behavior was analyzed according to Deacon *et al.*, a burrowing tube (15 × 0.03 × 6.5 cm) was filled with 200-g food pellets (ssniff-Spezialdiäten GmbH) and was placed in the cage 2.5–3 h before the dark phase and 1–1.5 h after the last injection of the chemotherapeutics (Gallo, CHC, or the appropriate vehicle) at 4:00–4:30 p.m. in the animal facility.[24,25] After 2 h, the amount of pellets displaced from the tube was calculated. Animals, which burrowed less than 100-g pellets during the preexperimental phase, were excluded from further analysis (20.8% of the animals). To analyze the nesting behavior, a nestlet was supplied for each mouse 1–2 h before the dark phase (6:00–6:30 p.m.) (5-cm square of pressed cotton batting, ZOONLAB GmbH, Castrop-Rauxel, Germany) and on the next day (9:00–11:00 a.m.), the score of the nest was assessed similar to a point scale from Deacon.[24] In addition to 1–5 points from Deacon, we scored 6 points for a perfect nest (>90% is torn) that looks like a crater and more than 90% of the circumference of the nest wall is higher than the body height of the mouse. At the beginning of the experiment, both behavior assays were performed two times in group housing, since it was suggested that mice can learn from each other.[24] After the learning period during the whole experiment, all the animals were housed individually.

The concentrations of corticosterone were measured in blood plasma and its metabolites in feces

Figure 1. Experimental procedure for the pancreatic carcinoma model. Laparotomy was executed, and tumor cells were injected into the murine pancreas on day 0. After a few recovery days, the chemotherapeutic treatments were conducted from day 4 until day 37. Met (125 mg/kg) and CHC (15 mg/kg) were i.p. injected daily (A), while Gallo (20 mg/kg) was i.p. injected three times a week (B). All distress parameters were assessed before (pre) and directly after operation (op) until recovery day 2. During chemotherapy, all parameters were quantified during the early, middle, and late phase of the therapy.

(FCMs). Feces (200–400 mg) were collected from home cages 24 h after intervention, dried for 4 h at 65 °C, and stored at –20 °C. Afterward, 50 mg of the dry feces was extracted with 1 mL 80% methanol for subsequent analysis using a 5α-pregnane-3β,11β,21-triol-20-one enzyme immunoassay.[13,14] For the quantification of corticosterone in blood plasma, we used an additional 10 mice because the blood sampling procedure leads to a stress response[26] and might influence burrowing and nesting activity, as well as FCM concentrations. Six mice of an average age of 16.7 (16.1–17.3) weeks (median/interquartile range) and the average BW of 28.2 (25–30) g (median/interquartile range) received Met + Gallo treatment and four mice of a mean age of 17.4 (14.5–17.6) weeks (median/interquartile range) and an average weight of 26.2 (24.9–28) g (median/interquartile range) were treated with the corresponding sham intervention. Blood was collected 30 min after chemotherapy by a retro-orbital puncture after 2–3 min anesthesia with isoflurane (5%). The fast sampling via isoflurane within 3 min is mandatory to exclude an influence of the sampling method to the corticosterone level.[26,27] The blood samples were centrifuged (1200 × *g* for 10 min) and plasma was stored at –20 °C. Plasma corticosterone concentrations were measured using an ELISA-Kit (DEV 9922, Demeditec Diagnostics GmbH, Erfurt, Germany) according to the manufacturer's instructions.

Data analysis

Data for Figures 2–4 are presented in the form of a line graph revealing the mean value ± standard deviation for parametric data with equal variance and median ± interquartile range for nonparametric data. For statistical evaluation, the data were analyzed with the SigmaPlot® 12.0 (Systat Software Inc., San Jose, CA) program. The characteristics of data were assessed by the Shapiro–Wilk and Levene median equal variance tests. All post-hoc tests were performed as suggested by this software. In the case of parametric data, significances of differences during the perioperative phase were evaluated by a one-way repeated-measures ANOVA, with pairwise comparison by the Holm–Sidak method. In the case of nonparametric data, differences were analyzed by ANOVA on ranks (the Kruskal–Wallis test), followed by the Student–Newman–Keuls method. Significances of differences during the therapy phases were analyzed by a two-way repeated-measures ANOVA (treatment and phase), with pairwise comparison by the Holm–Sidak method. Differences with $P \leq 0.05$ were considered to be significant.
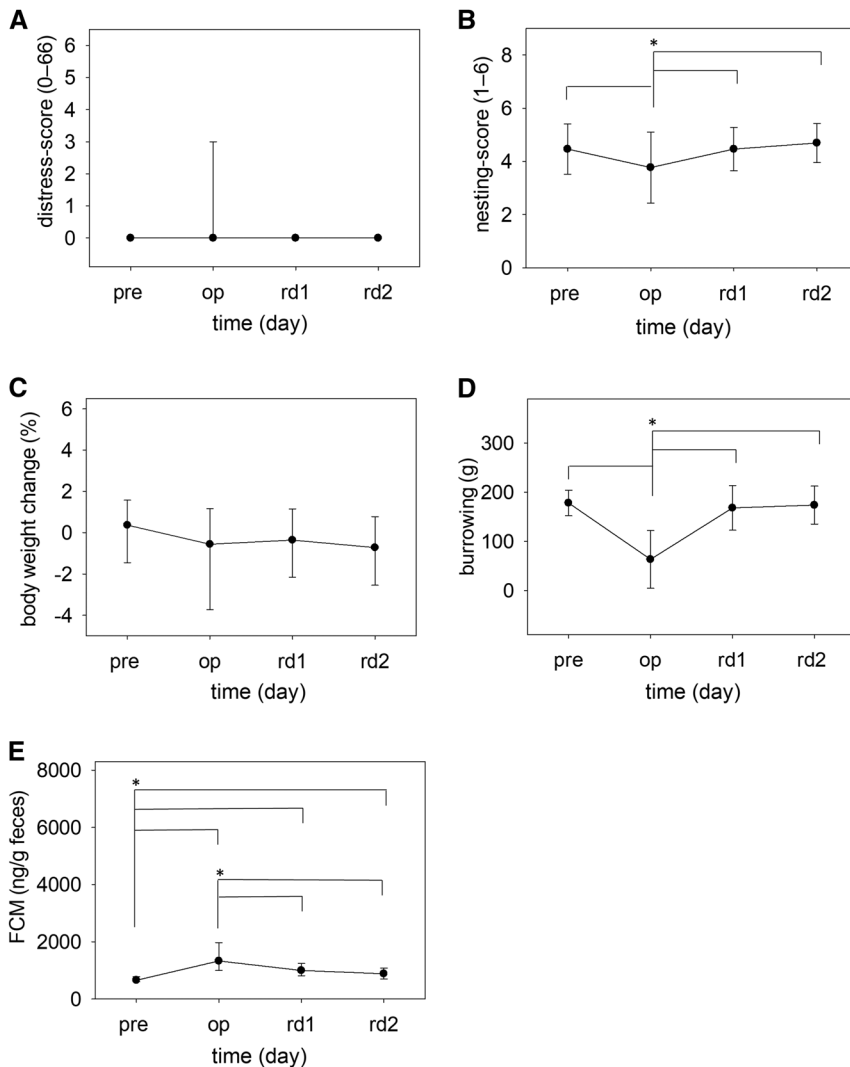
Figure 2. Animal distress after laparotomy and tumor cell injection. Distress score (A), nesting behavior (B), percentage of BW change (C), burrowing behavior (D), and FCMs (E) were evaluated before any intervention (pre), after operation (op), and during recovery (recovery days: rd1 and rd2). Parametric data with equal variance are described as the mean $\pm$ standard deviation, and significances were determined by a one-way repeated-measures ANOVA, followed by pairwise comparison with a Tukey test (B, D, and E). Nonparametric data are represented as the median $\pm$ interquartile range and were analyzed with a repeated-measures ANOVA on ranks (Kruskal–Wallis test), followed by the Student–Newman–Keuls method (A and C): *$P \leq 0.05$. (A, B, C, D, and E; $n = 26$).

To compare the performance of each parameter for distress quantification, receiver operating characteristic (ROC) curve analysis was performed. The data of all animals before and on the day of surgical intervention were used for distress prediction and the area under the ROC curve (AUC), the 95% confidence intervals (CIs), and the *P* value were calculated for each parameter. ROC curves are able

to graph the sensitivity of a diagnostic test.[28] The AUC of 1.0 means that the parameter is perfect to discriminate between the animals before and after operation, whereas a value of 0.5 indicates no discriminative power for this parameter. To analyze the AUC of the combination of all three parameters, the data sets were combined by multiple logistic regression model and the ROC curves were calculated
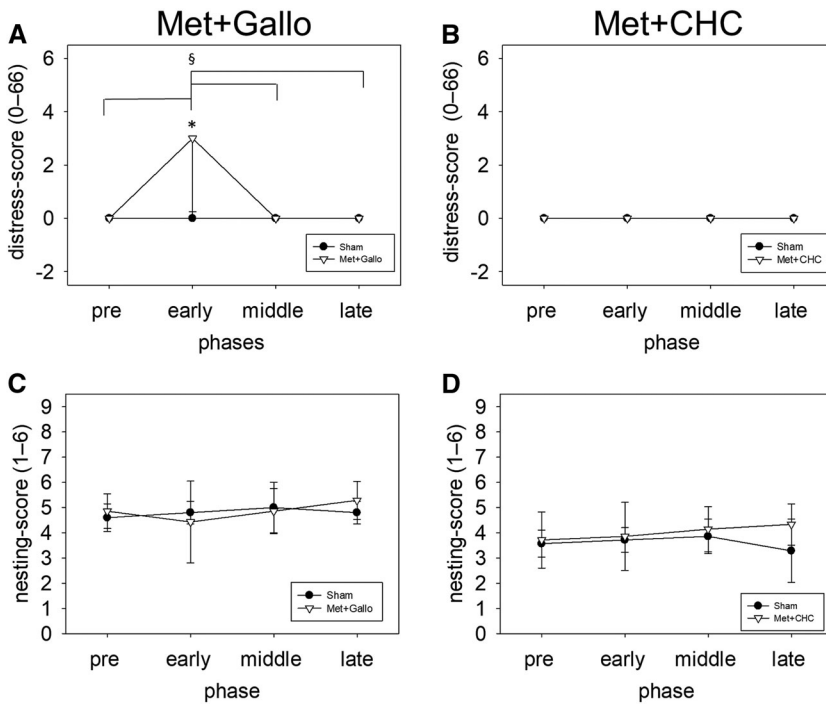
Figure 3. Animal distress during distinct chemotherapeutic treatments. The distress score was determined at the indicated phases after Met + Gallo treatment (A) or Met + CHC therapy (B), and compared with the distress score after injection of the appropriate vehicle (sham). Nesting behavior was evaluated after injection of the appropriate vehicle or after treatment with Met + Gallo (C) or Met + CHC (D). Nonparametric data are described as the median ± interquartile range (A and B), and parametric data with equal variance are represented as the mean ± SD (C and D). Significances were determined by a two-way repeated-measures ANOVA with pairwise comparison by the Holm–Sidak method: *$P < 0.05$, sham versus treatment groups; $^§P < 0.05$, phase versus phase within treatment groups (A + C: sham $n = 5$, and Met + Gallo, $n = 7$; B + D: sham, $n = 7$, and Met + CHC, $n = 7$).

afterward. The best cutoff for the single parameters, such as BW change, burrowing, and FCMs, was calculated by Youden's *J* index, and the data points during the therapy phases were distributed to distress levels 1 and 2, according to the cutoff values.

Multivariate model by logistic regression
Data collected from animals before (pre) and directly after surgery (op) ($n = 26$; 52 data point in total) were used as a training data set. The data table consisted of two independent variables (burrowing and FCMs) and the status of each data point (pre or postop) as the dependent variable. Both independent variables were tested for normality (the Shapiro–Wilk test), as well as multivariate normality (the Shapiro–Wilk multivariate normality test[29]) and in both cases, the tests failed to reject the null hypothesis for normality. Therefore, (multivariate)

logistic regression was used to fit the basic binomial model for the classification of sample states (pre/op, meaning severity level 1/2) in R.[30] Using the coefficients from the fit, the discriminator (line) was plotted into the two-dimensional space of the training data (variables $x =$ FCMs and $y =$ burrowing). The classification threshold for predictions was optimized using ROC analysis (cutoff at the combined sensitivity/specificity maximum).[30] The resulting threshold was then used in the subsequent predictions for both the training and test data to assess the model's performance and classification success. All data collected during the distinct therapy phases (early, middle, and late) were pooled as test data sets for each treatment (Met + Gallo, $n = 7$, 21 data points, or Met + α-cyano-4-hydroxycinnamate (Met + CHC) $n = 7$, 21 data points). Finally, these two categories of treatments were classified separately into either state (level 1 or 2).
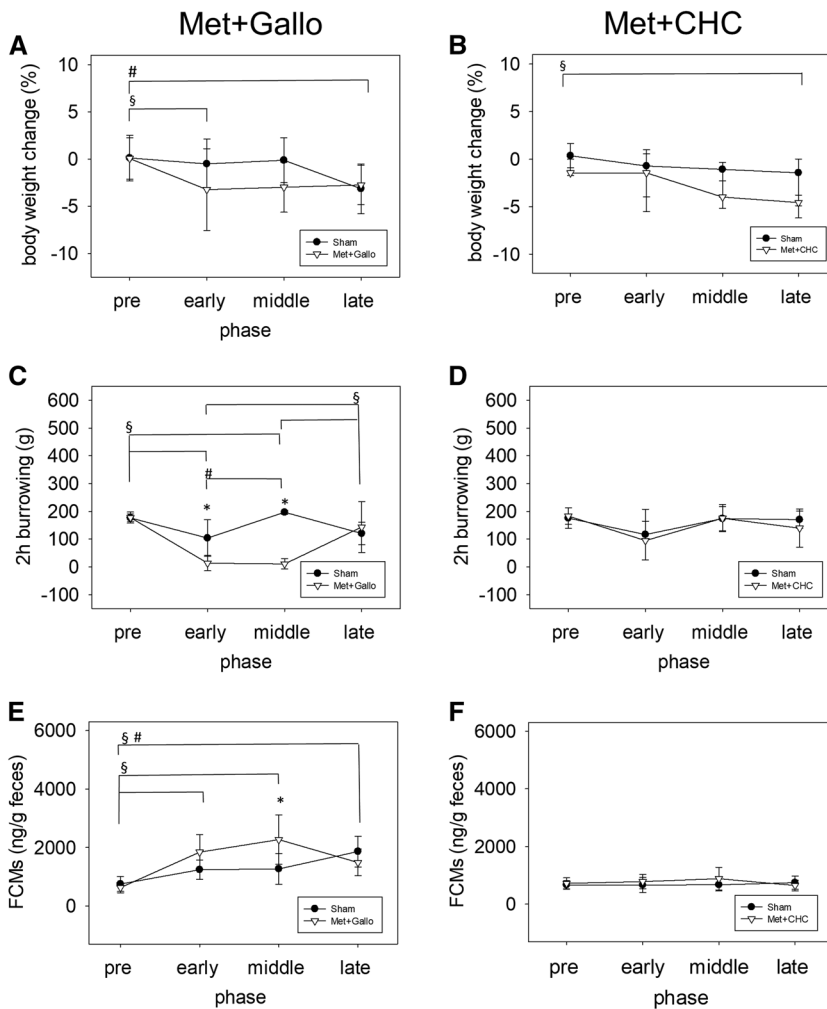
Figure 4. Analysis of sensitive distress parameters during distinct chemotherapeutic treatments. The percentage of BW change was analyzed after injection of Met + Gallo (A), Met + CHC (B), or the corresponding vehicle solutions (sham) during the indicated therapy phases. Burrowing behavior was quantified after Met + Gallo (C), Met + CHC (D), or sham treatment. FCM concentrations were analyzed after treatment with Met + Gallo (E), Met + CHC, or an appropriate vehicle (F). Parametric data with equal variance are described as the mean ± SD (A, C, D, E, and F), and nonparametric are illustrated as the median ± interquartile range (B). Significances were determined by a two-way repeated-measures ANOVA with pairwise comparisons by the Holm–Sidak method: *$P \leq 0.05$, sham versus treatment groups; [§]$P \leq 0.05$, phase versus phase within treatment group; [#]$P \leq 0.05$, phase versus phase within sham groups; (A, C, and E: sham, $n = 5$, and Met + Gallo, $n = 7$; B, D, and F: sham, $n = 7$, and Met + CHC, $n = 7$).

Results

The distress of laparotomy and tumor cell injection

For the evaluation of animal distress after laparotomy and tumor cell injection, all distress parameters were assessed before and directly after operation, and on recovery days 1 and 2 (Fig. 1). The distress score was slightly increased when assessed 30 min after operation owing to the fact that 42% of mice displayed an abnormal posture by a slightly curved position (score 3[20,21]). However, the mice recovered within 1 day (Fig. 2A). The nesting behavior was reduced significantly after surgery, followed by a fast recovery (Fig. 2B). We noticed a slightly nonsignificant reduction in BW after operation until recovery day 2 (Fig. 2C). The burrowing behavior was reduced significantly

after surgery but recovered within 1 day (Fig. 2D). We also observed a significant increase in FCM concentration after operation until recovery day 2 (Fig. 2E). When adding 40 more mice from other studies to assess distress score and BW change, a significant change after surgery can be observed for all distress parameters (Fig. S2, online only).

Distress during chemotherapeutic interventions

We evaluated the distress of two distinct chemotherapies, by treating mice either with (Met + Gallo) or (Met + CHC). These chemotherapeutic agents are reported to inhibit the metabolism of carcinoma cells.[31–33]

The distress score after Met + Gallo treatment was significantly increased during the early phase of therapy, reflected by a score of 3 (abnormal posture,[20,21] caused by stretching of the hind legs) in 57.1% of the mice. These symptoms were neither detectable in sham (Fig. 3A), nor in Met + CHC–treated mice (Fig. 3B). Mice of both combinatorial treatments constructed nearly perfect nests during the chemotherapeutic treatment, as given by a constant nesting score of 4 to 5 (Fig. 3C and D).

We observed a significant reduction in BW during the early therapy phase after Met + Gallo treatment, compared with the prevalues, while sham animals displayed only a significant BW reduction at the late phase of therapeutic intervention (Fig. 4A). After Met + CHC treatment, a significant BW loss was detected at the late therapy phase (Fig. 4B).

Burrowing behavior was significantly reduced after Met + Gallo treatment at the early and middle phases of therapeutic intervention (Fig. 4C). By contrast, after Met + CHC and its corresponding sham treatment, no significant reduction in burrowing behavior was observed throughout the entire therapy period (Fig. 4D).

After Met + Gallo treatment, FCM concentrations were increased throughout the therapy period, while sham-treated mice showed only a significant rise in corticosterone metabolites at the late phase of therapy (Fig. 4E). At the middle phase of therapy, significantly higher FCM concentrations were evaluated after Met + Gallo, compared with the sham group (Fig. 4E). We wanted to verify these data and assessed, therefore, the plasma corticosterone concentration on 10 additional mice. These results also indicated significant higher corticosterone

concentrations during the middle therapy phase (Fig. S3, online only). In contrast to Met + Gallo, the treatment with Met + CHC caused no increase in FCMs (Fig. 4F). In summary, Met + Gallo treatment seems to cause more distress to mice when compared with Met + CHC therapy. Especially, the parameters BW change, burrowing behavior, and FCMs proved to be sensitive for distress assessment after chemotherapy, owing to significant alterations in the therapy phases (Fig. 4A–F). However, few significant changes were observed in the assessment of distress–score and nesting behavior after both combinatorial treatments (Fig. 3). Therefore, we excluded these parameters when classifying distress during these chemotherapies.

Classification of distress during chemotherapies

In order to pursue a multifactorial distress analysis, we first evaluated the performance of BW change, burrowing behavior, and FCM concentration, when differentiating between distress measured before and after surgical intervention. For this purpose, we used ROC curve analysis. While the percentage of BW change displayed a low discriminative power ($A = 0.605$, 95% CI: 0.450–0.760) to differentiate between pre- and postoperative data, burrowing behavior ($A = 0.951$, 95% CI: 0.897 to –1.000) and FCMs ($A = 0.901$, 95% CI: 0.814–0.988) proved to be significantly (due to nonoverlapping CIs) better readout parameters (Fig. 5A–C). However, the combination of burrowing behavior and FCMs ($A = 0.966$, 95% CI: 0.918–1.000), as well as the combination of all three distress parameters ($A = 0.975$, 95% CI: 0.940–1.000), had the highest performance as indicated by the AUC (Fig. 5D and E).

With the goal to combine the most efficient readout parameters in a multifactorial distress analysis and to classify distinct severity levels, we generated a training model. Thus, the values of burrowing behavior and FCM concentrations were plotted for each animal before (pre) and after surgical intervention (op) in a 2D scatter diagram (Fig. 6A). Multivariate logistic regression separated two clusters by defining a discriminator (intercept (β_0) = –1.6368, $P = 0.5259$, burrowing (β_1) = 0.0467, $P = 0.0067$; FCMs (β_2) = –0.00572, $P = 0.0312$). The cluster, which included 96% data points of animals before any interventions, was
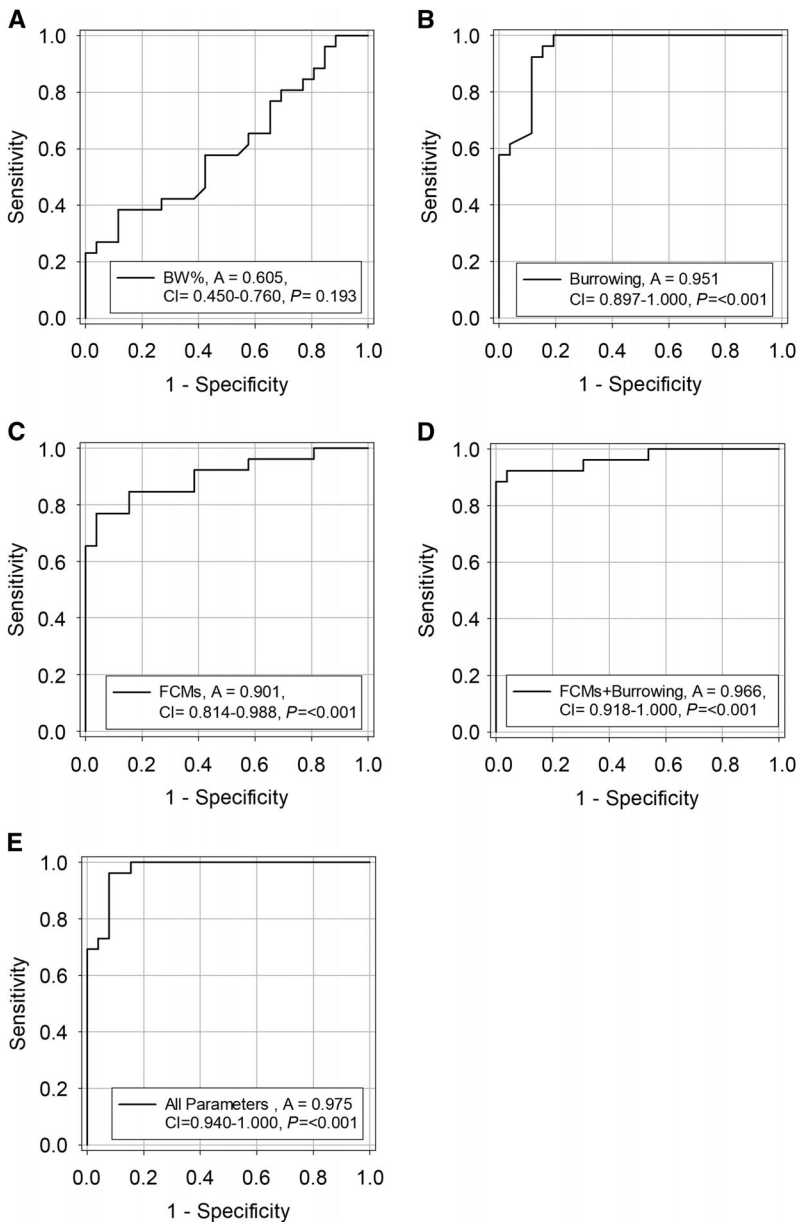
Figure 5. ROC curve analysis to evaluate the performance of each single parameter and the combination of parameters for distress quantification. Data before operation (pre) and after operation (op) from 26 mice were used to analyze the discriminatory power of each parameter. The AUC (A), 95% CI, and *P* value were assessed separately for BW changes (A), burrowing behavior (B), and FCMs (C). Additionally, these parameters were determined using multiple logistic regression for the combination of burrowing + FCMs (D) and the combination of all parameters (E). (A–E: pre and op, *n* = 26).

classified as severity level 1. The cluster consisting of 87% data points from animals after surgery was defined as distress level 2 (Fig. 6B). Using the established discriminator, to distinguish distress of animals before and after surgery, this training model achieved an accuracy of 0.9231, a sensitivity of 0.9583, and a specificity of 0.8926 (Fig. 6B).

To classify the distress caused during the therapy phases, the data assessed at the early, middle, and late phases after each therapy were pooled. The
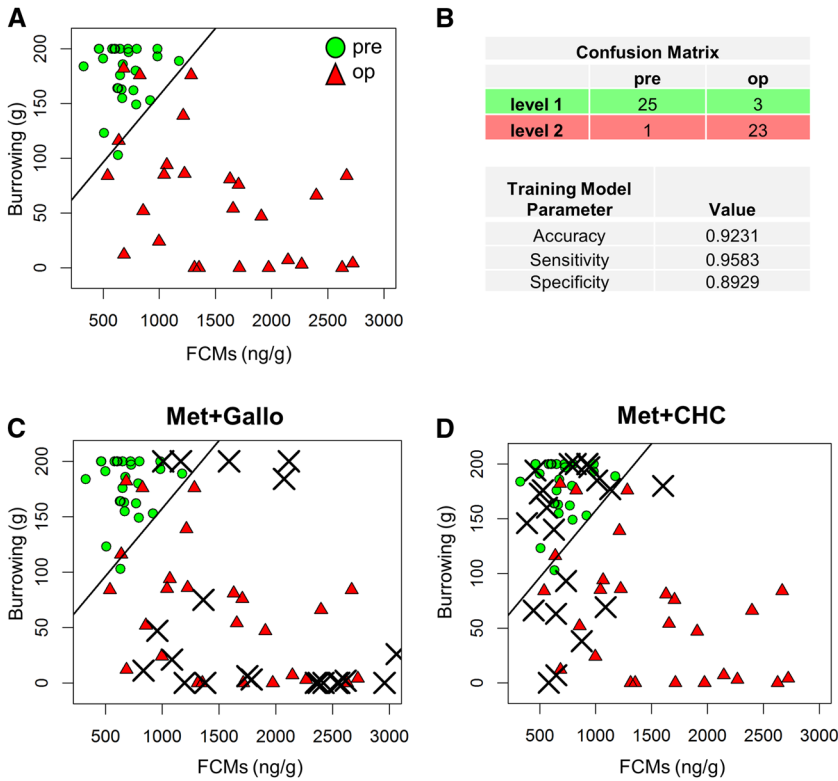
Figure 6. A generated 2D training model based on pre- and postoperative data was used to grade the severity of the distinct therapies. Single data points, which were derived from 26 animals before any intervention (pre; green dots) and after operation (op, red triangles), are presented in the form of a two-dimensional scatter plot, based on the parameters burrowing and FCMs (A). To generate a training model, these data points were spatially separated into two clusters using multivariate logistic regression. The cluster, including 96% data points of animals before any interventions, was classified as severity level 1 (B). The cluster consisting of 87% data points from animals after surgery was defined as distress level 2. Using the established discriminator, the model is further characterized by a confusion matrix showing the accuracy, sensitivity, and specificity of the training data (B). Single data points (crosses), which present the individual distress assessed from the identical animals during the therapy phases (early, middle, and late) after Met + Gallo (C) or Met + CHC intervention (D), were plotted in the training model. Pre, $n = 26$, and op, $n = 26$, 52 data points in total (A); Met + Gallo, $n = 7$, 21 data points (C), and Met + CHC, $n = 7$, 21 data points (D).

single test data points were plotted, respectively, for Met + Gallo and Met + CHC treatment into the generated training model and were separated into two distress levels by the discriminator (Fig. 6C and D). On this basis of this training model, two data points were assigned to distress level 1 and 19 data points to distress level 2 after Met + Gallo intervention (Table. 1). By contrast, after Met + CHC therapy, 12 data points were classified to distress level 1 and 9 data points to distress level 2. Thus, during Met + Gallo treatment, mice usually experienced a higher distress level 2, while animals mostly experienced a lower distress level 1 during Met + CHC treatment. Fisher's exact test confirmed a significant difference ($P = 0.003$) in the distress

Table 1. Distribution of distress level after Met + Gallo compared with Met + CHC intervention, according to the 2D scatter plot and multivariate logistic regression

	Met + Gallo	Met + CHC
Level 1	2	12
Level 2	19	9

NOTE: $P = 0.003$ by Fisher's exact test.

level distribution between these two therapies (Table 1). In addition, the efficiency of distress classification for single parameters, such as BW change, burrowing activity, and FCMs, was evaluated. For this purpose, the best cutoff between pre and

Table 2. Distribution of distress level by single variables after Met + Gallo and Met + CHC intervention, according to the calculated cutoff for the percentage of body weight change (−1.97%), burrowing (144 g), and FCMs (989 ng/g) by Youden's *J* index

	Body weight change (%)		Burrowing (g)		FCMs (ng/g)	
	Met + Gallo	Met + CHC	Met + Gallo	Met + CHC	Met + Gallo	Met + CHC
Level 1	6	7	5	13	2	17
Level 2	15	14	16	8	19	4
	nonsignificant		$P = 0.028$ by Fisher's exact test		$P < 0.001$ by Fisher's exact test	

postoperative (op) data was calculated by Youden's *J* index. Data assessed at the early, middle, and late phases after each therapy were pooled and assigned to distress level 1 or 2 according to the calculated cutoff. The cutoff for BW change was determined at −1.97%, for burrowing activity at 144 g, and for FCMs at 989.9 nanograms per gram. When assigning BW data according to the determined cutoff of −1.97%, no significant difference was observed when comparing Met + Gallo with Met + CHC therapy (Table 2). This is consistent with a low performance of BW when differentiating between pre and postoperative distress levels (Fig. 5A). When analyzing burrowing activity as a single readout parameter, animals experienced at most time points distress level 2 during Met + Gallo treatment, while animals mostly experienced distress level 1 during Met + CHC treatment (Table 2). Fisher's exact test confirmed a significant difference ($P = 0.028$) in the distress level distribution between these two therapies. When evaluating FCMs, animals mostly experienced distress level 2 during Met + Gallo, while Met + CHC–treated animals usually experienced distress level 1 (Table 2). Fisher's exact test confirmed a significant difference ($P < 0.001$) in the distribution between these two therapies. These results demonstrate that Met + Gallo treatment causes more distress than Met + CHC intervention.

Discussion

This study concluded that simple, noninvasive methods can identify the side effects of therapies and grade the severity of animal experiments. A 2D scatter plot followed by multivariate logistic regression was able to combine multiple readout parameters when defining specific distress levels. Similar distress levels were also obtained, when using a single readout parameter, by the application of ROC curve analysis and Youden's *J* index. Such

multi and univariate methods might be beneficial for future evaluation of animal experiments and might detect side effects of therapies.

Multivariate analysis was also applied in other studies to analyze the distress of animals. For example, Häger *et al.* assessed the severity of a murine colitis model and restrain stress intervention by *k*-means clustering of BW and voluntary wheel running data.[34] Seiffert *et al.* used principal component analysis to quantify, if tethered or telemetric monitoring of epileptic seizures causes more distress to rats.[35] In contrast to these studies, we critically evaluated the efficiency of single parameters for distress quantification compared with a multifactorial method. Using burrowing and FCMs as a single readout parameter, as well as using them in the form of a multifactorial analysis, supported the same conclusion. Thus, one could critically ask the question, whether it is not enough to rely on a single parameter. However, one of our previous studies demonstrated that the performance of different readout parameters differs between distinct animal models, such as chronic pancreatitis or laparotomy.[31] In addition, some of the used parameters might not score exclusively distress. They might also be influenced by positive excitement or different physiological responses.[6] Even corticosterone is known to be influenced by, for example, the circadian rhythm,[36–38] estrus circle,[36,39] or sexual arousal.[40] Thus, relying on only one parameter increases the risk to reach an incorrect conclusion. This potential bias is reduced when using a multivariate distress analysis. However, the subjective bias during the assessment of the parameters was unfortunately not reduced by randomization or blinding, which is a limitation of the present study.

In this study, we determined two distinct severity levels using the univariate and multivariate analysis. Severity level 1 was defined in the 2D scatter plot

by a cluster, which included 96% of the data points measured before any intervention (Fig. 6). The univariate discrimination of burrowing behavior and FCMs by Youden's *J* index obtained a similar confusion matrix with 91.7–96.0% of data points assessed before intervention assigned to distress level 1 (data not shown). The burrowing activity, as well as the FCM concentrations, is similar to data of healthy mice[41,42] or mice undergoing very mild stressors, such as single isoflurane anesthesia.[43] We, therefore, suggest that severity level 1 represents "mild" distress.

Severity level 2 was defined as a cluster that included 87% of the data points after and 4% of data points before surgery (Fig. 6). A similar reduction of burrowing activity (∼50–100%) or an increase in FCM concentrations (3- to 4-fold) could also be observed after intrabone marrow transplantation[44] or during chronic pancreatitis.[18] The EU-Directive 2010/63/EU suggests that surgical interventions with proper anesthesia might be ranked as moderate severity.[4] According to the fact that the cluster defining severity level 2 obtained most data measured after surgical intervention and in agreement with the above-cited literature and considering the EU-Directive 2010/63/EU, we suggest that the severity level 2 represents "moderate" distress. We recognized that not only surgery but also anesthesia and analgesia may influence our used distress parameters. For example, isoflurane anesthesia without any surgery is reported to provoke significant alterations of burrowing and nesting activity.[10,45,46] However, anesthesia and analgesia are mandatory for surgical interventions and we did not separately analyze these aspects in this study.

On the basis of the classification of all data points to distress levels 1 and 2, we can define that mice after the Met + Gallo intervention experience mainly "moderate" and rarely "mild" distress. By contrast, the distress caused by the Met + CHC intervention can be mainly classified as "mild." These data are consistent with clinical studies, which describe that different chemotherapeutic treatments have distinct side effects and can, therefore, cause different levels of distress.[47] Met is known as an antidiabetic drug with recently described anticancer effects.[48] Therefore, Met is used in several clinical trials as a treatment for different cancer types, with minor side effects.[49] Gallo and CHC were not used in clinical trials so far and

little is known about potential side effects. During the early phase of Gallo treatment, we observed that some mice showed writhing behavior, which is expressed by a stretching of the hind legs and a pressing of the abdomen into the substrate 30 min after injection. This behavior is known to be a sign of abdominal discomfort[50] and supports the assumption that Gallo causes drug-specific distress in animals. Writhing behavior is mostly associated with visceral pain,[50] which might imply that the used analgesic metamizol is unable to completely cover the pain induced by Gallo. Thus, the concentration of the self-administered metamizol might be insufficient in some mice, or even more efficient analgesics, such as opioids, might be required in future animal studies to cover drug-induced pain. However, in the late phase of chemotherapy, these animals seem to desensitize to the Gallo injection, showing fewer signs of discomfort (Fig. 3A). This observation is in line with reports that mice are able to habituate to chronic stressors like intraperitoneal (i.p.) injections.[37] Thus, our results indicate that besides severity assessment, multivariate or univariate distress analysis could also be used for judging side effects of therapies. Toxicological analysis of potent drugs in animals is mandatory before clinical trials. Current toxicological studies include monitoring of the cardiovascular and respiratory system combined with histopathological, biochemical, immunological, and hematological analysis.[51–53] Blood collection and analytic methods to monitor the cardiovascular and respiratory systems are highly invasive or require the restraining of animals and might, therefore, influence physiological readout parameters.[52] However, using noninvasive parameters, such as burrowing and FCMs, in longitudinal studies allows to detect side effects of some therapies without additional stress induction. This would, therefore, contribute to the refinement of pharmaceutical research without increasing the number of animals. For example, cardiovascular monitoring of rodents is mainly performed with implanted radiotelemetric transmitters.[54] Our working group analyzed in a previous study that noninvasive parameters, such as burrowing, nesting, FCMs, and BW, are as efficient as continuous monitoring of heart rate, body temperature, and activity by telemetry for distress analysis.[17]

Burrowing activity is also comparable with the so-called "activities of daily living" in humans

and impairment can be an early indicator for neurological disorders.[55–57] Activity and motoric dysfunction of rodents are currently used to identify the potential neurotoxicity of drugs.[58] The addition of noninvasive behavior tests, such as burrowing activity, might be the first step to assess side effects, such as headache, dizziness, or mental disturbances, which can barely be assessed by traditional toxicological methods.[8]

Besides contributing to the refinement by evaluating the side effects of drugs, the presented multivariate model can be used as a practical tool to assess distress in many different animal models and interventions. This could provide the basis not only for correctly classifying animal models according to their severity but could also be used to assess the effect of refinement methods in animal-based research, that is, analgesia or environmental enrichment.

Acknowledgments

Author contributions

D.Z. and B.V. acquired the funding for this project. D.Z., B.V., and S.K. designed the project. S.K., E.H.U.W., J.E., M.S., D.Z., R.P., and X.Z. acquired the data. S.K., S.R.T., D.Z., A.A, C.H., and A.B. did the statistical analysis. S.K., B.V., D.Z., and S.R.T. wrote the manuscript. All authors approved the final version of the manuscript.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1. Flowchart: sample size determination for experiments.

Figure S2. Animal distress after laparotomy and tumor cell injection, using up to 40 additional mice.

Figure S3. Corticosterone concentration in blood plasma.

Competing interests

The authors declare no competing interests.

References

1. Prescott, M.J. & K. Lidster. 2017. Improving quality of science through better animal welfare: the NC3Rs strategy. *Lab. Anim. (NY)* **46:** 152–156.

2. Graham, M.L. & M.J. Prescott. 2015. The multifactorial role of the 3Rs in shifting the harm–benefit analysis in animal models of disease. *Eur. J. Pharmacol.* **759:** 19–29.

3. European Commission. 2013. Caring for animals, aiming for better science. Accessed on January 20, 2020. https://ec.europa.eu/environment/chemicals/lab_animals/pdf/guidance/project_evaluation/en.pdf.

4. The European Parliament and the Council of the European Union. 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. 2010/63/EU.

5. Smith, D., D. Anderson, A.-D. Degryse, *et al.* 2018. Classification and reporting of severity experienced by animals used in scientific procedures: FELASA/ECLAM/ESLAV Working Group report. *Lab. Anim.* **52:** 5–57.

6. Hawkins, P., D.B. Morton, O. Burman, *et al.* 2011. A guide to defining and implementing protocols for the welfare assessment of laboratory animals: eleventh report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Lab. Anim.* **45:** 1–13.

7. Mellor, D.J. & C.S.W. Reid. 1994. *Concepts of animal well-being and predicting the impact of procedures on experimental animals. Improving the well-being of animals in research environment.* Australian and New Zealand Council for Care of Animals in Research and Teaching (ANZCCART). 3–18.

8. 2004. *Intentional human dosing studies for EPA regulatory purposes: scientific and ethical issues.* Washington, DC: The National Academies Press.

9. Jirkof, P., N. Cesarovic, A. Rettich, *et al.* 2010. Burrowing behavior as an indicator of post-laparotomy pain in mice. *Front. Behav. Neurosci.* **4:** 165.

10. Jirkof, P., T. Fleischmann, N. Cesarovic, *et al.* 2013. Assessment of postsurgical distress and pain in laboratory mice by nest complexity scoring. *Lab. Anim.* **47:** 153–161.

11. Peng, M., C. Zhang, Y. Dong, *et al.* 2016. Battery of behavioral tests in mice to study postoperative delirium. *Sci. Rep.* **6:** 29874.

12. Benedetti, M., R. Merino, R. Kusuda, *et al.* 2012. Plasma corticosterone levels in mouse models of pain. *Eur. J. Pain* **16:** 803–815.

13. Touma, C., N. Sachser, E. Möstl, *et al.* 2003. Effects of sex and time of day on metabolism and excretion of corticosterone in urine and feces of mice. *Gen. Comp. Endocrinol.* **130:** 267–278.

14. Touma, C., R. Palme & N. Sachser. 2004. Analyzing corticosterone metabolites in fecal samples of mice: a noninvasive technique to monitor stress hormones. *Horm. Behav.* **45:** 10–22.

15. Palme, R. 2019. Non-invasive measurement of glucocorticoids. Advances and problems. *Physiol. Behav.* **199:** 229–243.

16. Kilkenny, C., W.J. Browne, I.C. Cuthill, *et al.* 2010. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol.* **8:** e1000412.

17. Kumstel, S., P. Vasudevan, R. Palme, *et al.* 2020. Benefits of non-invasive methods compared to telemetry for distress analysis in a murine model of pancreatic cancer. *J. Adv. Res.* **21:** 35–47.

18. Abdelrahman, A., S. Kumstel, X. Zhang, *et al.* 2019. A novel multi-parametric analysis of non-invasive methods to assess animal distress during chronic pancreatitis. *Sci. Rep.* **9:** 14084.

19. Stumpf, F., H. Algül, C.K. Thoeringer, *et al.* 2016. Metamizol relieves pain without interfering with cerulein-induced acute pancreatitis in mice. *Pancreas* **45:** 572–578.

20. Kumstel, S., G. Tang, X. Zhang, *et al.* 2019. Grading distress of different animal models for gastrointestinal diseases based on plasma corticosterone kinetics. *Animals (Basel)* **9:** E145.

21. Zhang, X., S. Kumstel, G. Tang, *et al.* 2019. A rational approach of early humane endpoint determination in a murine model for cholestasis. *ALTEX* https://doi.org/10.14573/altex.1909111.

22. Morton, D. & P. Griffiths. 1985. Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *Vet. Rec.* **116:** 431–436.

23. Paster, E.V., K.A. Villines & D.L. Hickman. 2009. Endpoints for mouse abdominal tumor models. Refinement of current criteria. *Comp. Med.* **59:** 234–241.

24. Deacon, R. 2012. Assessing burrowing, nest construction, and hoarding in mice. *J. Vis. Exp.* e2607. https://doi.org/10.3791/2607.

25. Deacon, R.M.J. 2006. Burrowing in rodents: a sensitive method for detecting behavioral dysfunction. *Nat. Protoc.* **1:** 118–121.

26. van Herck, H., V. Baumans, C.J. Brandt, *et al.* 2001. Blood sampling from the retro-orbital plexus, the saphenous vein and the tail vein in rats: comparative effects on selected behavioural and blood variables. *Lab. Anim.* **35:** 131–139.

27. Vahl, T.P., Y.M. Ulrich-Lai, M.M. Ostrander, *et al.* 2005. Comparative analysis of ACTH and corticosterone sampling methods in rats. *Am. J. Physiol. Endocrinol. Metab.* **289:** E823–E828.

28. Hajian-Tilaki, K. 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J. Intern. Med.* **4:** 627–635.

29. Slawomir, J. 2012. mvnormtest: normality test for multivariables. R package version 0.1-9. Accessed on October 18, 2019. https://CRAN.R-project.org/package=MVnormtest.

30. R Core Team. 2019. R: a language and environment for statistical computing. Accessed on October 18, 2019. https://www.R-project.org/.

31. Foretz, M., B. Guigas, L. Bertrand, *et al.* 2014. Metformin. From mechanisms of action to therapies. *Cell Metab.* **20:** 953–966.

32. Colen, C.B., Y. Shen, F. Ghoddoussi, *et al.* 2011. Metabolic targeting of lactate efflux by malignant glioma inhibits invasiveness and induces necrosis. An *in vivo* study. *Neoplasia* **13:** 620–632.

33. Farabegoli, F., M. Vettraino, M. Manerba, *et al.* 2012. Galloflavin, a new lactate dehydrogenase inhibitor, induces the death of human breast cancer cells with different glycolytic attitude by affecting distinct signaling pathways. *Eur. J. Pharm. Sci.* **47:** 729–738.

34. Häger, C., L.M. Keubler, S.R. Talbot, *et al.* 2018. Running in the wheel: defining individual severity levels in mice. *PLoS Biol.* **16:** e2006159.

35. Seiffert, I., R.M. van Dijk, I. Koska, *et al.* 2019. Toward evidence-based severity assessment in rat models with repeated seizures: III. Electrical post-status epilepticus model. *Epilepsia* **60:** 1539–1551.

36. Gong, S., Y.-L. Miao, G.-Z. Jiao, *et al.* 2015. Dynamics and correlation of serum cortisol and corticosterone under different physiological or stressful conditions in mice. *PLoS One* **10:** e0117503.

37. Dalm, S., L. Enthoven, O.C. Meijer, *et al.* 2005. Age-related changes in hypothalamic—pituitary–adrenal axis activity of male C57BL/6J mice. *Neuroendocrinology* **81:** 372–380.

38. Teilmann, A.C., O. Kalliokoski, D.B. Sørensen, *et al.* 2014. Manual versus automated blood sampling. Impact of repeated blood sampling on stress parameters and behavior in male NMRI mice. *Lab. Anim.* **48:** 278–291.

39. Nichols, D.J. & P.F. Chevins. 1981. Plasma corticosterone fluctuations during the oestrous cycle of the house mouse. *Experientia* **37:** 319–320.

40. Bronson, F.H. & C. Desjardins. 1982. Endocrine responses to sexual arousal in male mice. *Endocrinology* **111:** 1286–1291.

41. Gurfein, B.T., B. Hasdemir, J.M. Milush, *et al.* 2017. Enriched environment and stress exposure influence splenic B lymphocyte composition. *PLoS One* **12:** e0180771.

42. Möller, C., F. Wolf, R.M. van Dijk, *et al.* 2018. Toward evidence-based severity assessment in rat models with repeated seizures. I. Electrical kindling. *Epilepsia* **59:** 765–777.

43. Hohlbaum, K., B. Bert, S. Dietze, *et al.* 2017. Severity classification of repeated isoflurane anesthesia in C57BL/6JRj mice—assessing the degree of distress. *PLoS One* **12:** e0179588.

44. Pfeiffenberger, U., T. Yau, D. Fink, *et al.* 2015. Assessment and refinement of intra-bone marrow transplantation in mice. *Lab. Anim.* **49:** 121–131.

45. Cesarovic, N., F. Nicholls, A. Rettich, *et al.* 2010. Isoflurane and sevoflurane provide equally effective anaesthesia in laboratory mice. *Lab. Anim.* **44:** 329–336.

46. Jirkof, P., N. Cesarovic, A. Rettich, *et al.* 2012. Individual housing of female mice: influence on postsurgical behaviour and recovery. *Lab. Anim.* **46:** 325–334.

47. Partridge, A.H., H.J. Burstein & E.P. Winer. 2001. Side effects of chemotherapy and combined chemohormonal therapy in women with early-stage breast cancer. *J. Natl. Cancer Inst. Monogr.* **2001:** 135–142.

48. Decensi, A., M. Puntoni, P. Goodwin, *et al.* 2010. Metformin and cancer risk in diabetic patients. A systematic review and meta-analysis. *Cancer Prev. Res. (Phila.)* **3:** 1451–1461.

49. Chae, Y.K., A. Arya, M.-K. Malecek, *et al.* 2016. Repurposing metformin for cancer treatment. Current clinical studies. *Oncotarget* **7:** 40767–40780.

50. Ness, T.J. 1999. Models of visceral nociception. *ILAR J.* **40:** 119–128.

51. Greaves, P., A. Williams & M. Eve. 2004. First dose of potential new medicines to humans: how animals help. *Nat. Rev. Drug Discov.* **3:** 226–236.

52. Luft, J. & G. Bode. 2002. Integration of safety pharmacology endpoints into toxicology studies. *Fund. Clin. Pharmacol.* **16:** 91–103.

53. Rix, A., N.I. Drude, A. Mrugalla, *et al.* 2019. Assessment of chemotherapy-induced organ damage with Ga-68 labeled duramycin. *Mol. Imaging Biol.* https://doi.org/10.1007/s11307-019-01417-3.

54. Farraj, A.K., M.S. Hazari & W.E. Cascio. 2011. The utility of the small rodent electrocardiogram in toxicology. *Toxicol. Sci.* **121:** 11–30.

55. Deacon, R.M.J., D. Reisel, V.H. Perry, *et al.* 2005. Hippocampal scrapie infection impairs operant DRL performance in mice. *Behav. Brain Res.* **157:** 99–105.

56. Felton, L.M., C. Cunningham, E.L. Rankine, *et al.* 2005. MCP-1 and murine prion disease: separation of early behavioural dysfunction from overt clinical disease. *Neurobiol. Dis.* **20:** 283–295.

57. Yuan, D., C. Liu, J. Wu, *et al.* 2018. Nest-building activity as a reproducible and long-term stroke deficit test in a mouse model of stroke. *Brain Behav.* **8:** e00993.

58. United States Environmental Protection Agency (EPA). 1996. Health effects test guidelines OPPTS 870.6200 neurotoxicity screening battery public draft. Accessed on November 10, 2019. https://nepis.epa.gov.