Evaluating distress of animals in mouse models for liver fibrosis

Cumulative dissertation

To obtain the academic degree

Doctor of Medicine (Dr. med.)

From the Rudolf-Zenker-Institute for Experimental Surgery

Rostock University Medical Center

Director: Prof. Dr. med. Brigitte Vollmar

Submitted by Guanglin Tang

Born on 09/01/1988, Liaoning, China

Rostock, 2023

https://doi.org/10.18453/rosdok_id00005031

Reviewers

Reviewer 1:

Apl-Prof. Dietmar Zechner (Ph.D. in Biology, UCSD),

Universitätsmedizin Rostock

Reviewer 2:

Apl-Prof. Dr. med. Robert Jaster,

Universitätsmedizin Rostock

Reviewer 3:

Prof. Dr. Kerstin Schwabe,
Medizinische Hochschule Hannover

Date of submission: 02.01.2024

Date of defense: 02.07.2024

Studies of the cumulative dissertation

Study I:
Kumstel S, **Tang G**, Zhang X, Kerndl H, Vollmar B, Zechner D. Grading Distress of Different Animal Models for Gastrointestinal Diseases Based on Plasma Corticosterone Kinetics. Animals (Basel). 2019 Apr 3;9(4):145. (IF 2022: 3.0)

Study II:
Tang G, Seume N, Häger C, Kumstel S, Abshagen K, Bleich A, Vollmar B, Talbot SR, Zhang X, Zechner D. Comparing distress of mouse models for liver damage. Sci Rep. 2020 Nov 13;10(1):19814. (IF 2022: 4.6)

Study III:
Zhang X[#], Kumstel S[#], **Tang G[#]**, Talbot SR, Seume N, Abshagen K, Vollmar B, Zechner D. A rational approach of early humane endpoint determination in a murine model for cholestasis. ALTEX. 2020;37(2):197-207. (IF 2022: 5.6)
[#]equal contribution

Table of Contents

1. Abstract

High animal welfare standards are not only a prerequisite to get the permission to perform animal based research, but also provide an important foundation for high-quality biomedical research. In the last decades, many animal models of liver fibrosis have been developed for medical research. However, studies that define which methods are appropriate to measure severity of suffering, compare distress between animal models, or define early humane endpoints in laboratory animals remain scarce. In the first study, we analyzed the peak and duration of corticosterone kinetics, showing that both the absolute value and duration of corticosterone concentration are important readout parameters for evaluating distress in animals. In the second study, ROC curves were used, in order to analyse the diagnostic power of a distress score, burrowing activity and body weight change in an animal model for liver fibrosis. In addition, Support Vector Machine classification was carried out to compare the distress caused by BDL to the distress of repetitive CCl_4 intoxication. This approach demonstrated that BDL causes much more distress than CCl_4 induced liver damage. In the third study, Cox proportional-hazards model and Harrell's concordance index were used to analyze the performance of individual and multiple animal welfare parameters. A weight loss of 10-20%, combined with a reduction in burrowing activity of more than 79.4%, predicted death of these animals within two days. This could serve as an early humane endpoint and might reduce suffering of animals in future experiments. In conclusion, these results can provide an objective basis for minimizing animal suffering and improving the experimental quality of biomedical research.

2. Introduction

2.1 Improving quality of medical research through high animal welfare standard

Animal experiments have made important contributions to the development of medicine and human health. Many important discoveries in clinical medicine have come from animal research, and countless lives have been improved or saved by animal experiments. For example, animals have been used to develop vaccines to prevent diseases that previously killed billions of people, including smallpox, cholera, and polio.[1-3] The research results obtained by replicating human disease models in animals through drugs, surgery and other methods are important for us to understand different diseases, as well as design and test new disease treatments.[4, 5] This is the first step in medical discovery. From 1901 to 2021, 188 of the 225 Nobel Prize recipients in the physiological or medical category used animal models in their research, accounting for 83% of the total number of laureates.[6] Thus, experimental animals deserve a high standard in animal welfare. According to animal protection laws enacted by most nations,[7, 8] high animal welfare standards are also a prerequisite to get the permission to perform animal based research. In addition, high-level animal welfare also provide an important basis for high-quality biomedical research.[9, 10] Distress, pain and suffering can cause behavioral, physiological, and immunological changes in laboratory animals, which do not only lead to a decline in quality of life, but ultimately affects the reliability and repeatability of experiments. It is in the common interest of scientists and the public to make every effort to minimize the suffering of laboratory animals without interfering with research objectives. For each species and disease model, identifying appropriate indicators of distress to assess severity of suffering and determine humane endpoints are prerequisites for improving animal welfare.

2.2 Methods for distress evaluation

Mice are the main species used as laboratory animal. In 2017, 66% of all animals used for scientific purposes in Germany were mice.[11] Therefore, it is important to focus on evaluating distress in this rodent. A good selection of sensitive scientific methods to measure distress is not only the basis for evaluating the severity of animal models, but also the premise of rational selection of animal models and formulation of improvement strategies. The severity assessment of mice mainly included subjective and objective assessments. Subjective assessment refers to the subjective judgment of individual scientists based on the distress score sheet or a nesting score. Objective assessments were made primarily by measuring spontaneous behavior (such as burrowing activity)[12], physiology (such as heart rate and body weight),[13, 14] and stress hormones by biochemical assays.[15-17]

A welfare assessment protocol was established by Morton and Griffiths in 1985.[18] This score sheet included criteria such as dehydration to identify the animal's pain, suffering or distress

in a particular process.[18] For different animal models, many different scoring systems have been established to assess the severity of animal suffering.[19] Common criteria include weight loss, appearance, spontaneous and escape behavior, and specific clinical symptoms of interventions.[20] However many aspects of such distress scores such as apathy, are mainly based on the subjective observation of the researcher. This may cause incorrect conclusions. Thus, often a combination of distress scores plus other measures of distress is used to evaluate suffering of animals.

Nesting or burrowing behaviors are common in rodents. Wild house mice maintain heat by building nests, which also helps to avoid predators and competitors. The motivation and ability to build a nest still exist in domesticated experimental mice. Burrowing behavior refers to the typical behavior of mice that spontaneously remove items from the tubes in their home cages. The mice use the burrowing behavior to prepare for defense and escape. Burrowing was first measured by Deacon et al., [21] and was subsequently used to measure animal welfare by assessing the weight of pellets that were displaced from a tube.[22]

Body weight is a basic indicator for reflecting the physiological status of experimental animals and a sensitive indicator for the severity of suffering. A Reduction of body weight is mainly caused by reduced energy intake (such as indigestion or loss of appetite) and increased energy expenditure in laboratory animals. Many other studies have reported that body weight is very useful for determining human endpoints.[23-25]

Corticosterone, the main glucocorticoid regulating stress response in mice, has been demonstrated to be a sensitive indicator of welfare.[11, 26] Difference stressors induce the activation of the hypothalamus-pituitary-adrenal (HPA) axis and stimulate the hypothalamus to secrete corticotrophin-releasing hormone (CRH). CRH then leads to the secretion of adrenocorticotrophic hormone (ACTH) in the pituitary gland, which stimulates the adrenal cortex to release corticosterone into circulation. Corticosterone provides energy supply for stress response by regulating the metabolism and the defense response to stress.[16, 27]

2.3 Early humane endpoint determination for timely euthanasia

Determining a humane endpoint for timely euthanasia is a key aspect of reducing animal suffering. However, current criteria for humane endpoints tend to reflect severe suffering before death, such as 20% weight loss[28, 29] and hypothermia or lethargy.[30, 31] Defining more "humane" criteria, which can predict death within a period of time, could reduce suffering in laboratory animals. Common criteria to measure distress include body weight loss, appearance, spontaneous and flight behavior, and specific clinical symptoms of the intervention.[18-20] In addition, the rodents' natural behaviors, such as burrowing and nesting activities, were used to analyze health status of the animals.[32-35] This approach proved to be very useful for detecting abdominal and post-operative pain or stress in rodents.[22, 34, 36-38]

2.4 Animal models for liver fibrosis

Liver fibrosis is caused by many chronic liver diseases and often proceeds to liver cirrhosis and failure.[39, 40] The result of progressive fibrosis, cirrhosis, causes an estimated 1.16 million deaths worldwide each year, accounting for 2.4% of all deaths.[41, 42] Animal models for liver fibrosis are necessary in order to understand the pathogenesis of fibrosis and to explore potential therapies.

Several mouse models of liver fibrosis, for example, induced by hepatotoxin or bile duct ligation (BDL) have been established (see Figure 1).[43, 44] BDL causes cholestasis induced fibrosis, which may occur in humans with bile duct obstruction (cholelithiasis and tumour compression of bile ducts) and autoimmune diseases (primary biliary cirrhosis and primary sclerosing cirrhosis).[45] Repetitive administration of the hepatotoxin carbon tetrachloride (CCl_4) mimics liver damage in human by different toxins, such as drugs and alcohol.[43] Although mouse models of liver fibrosis are used more and more widespread in medical research, there are few available reports on distress evaluation in these models. Thus, studies defining and exploring how to reduce distress are necessary in this scientific field to improve animal welfare.

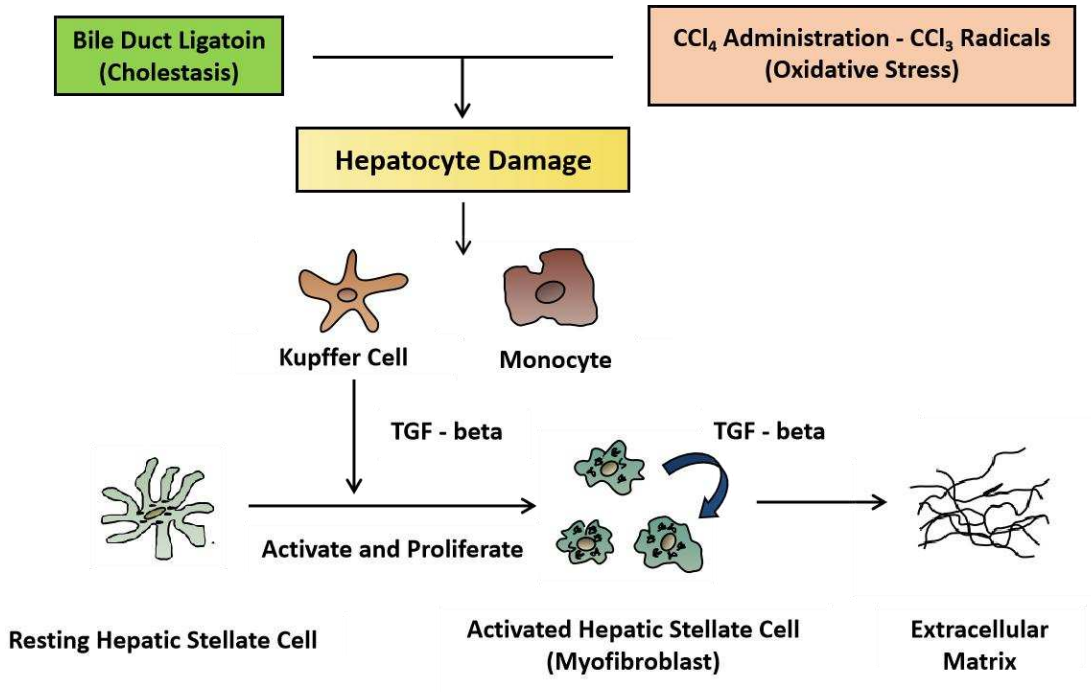


Figure 1. Pathogenesis of liver fibrosis.

2.5 Aim of this dissertation

It was the aim of our studies

- to compare the corticosterone kinetics between two distinct methods for damaging the liver,
- to explore the diagnostic power of different non-invasive methods to detect distress,
- to apply multivariate analysis combining multiple parameters for distress assessment, and
- to explore a reasonable method to determine early human endpoints for bile duct ligated mice.

It was the overarching objective of our studies to provide a solid basis for reducing animal suffering in future experiments and to help to improve preclinical research.

3. Material and methods

All methods were briefly described in the present cumulative dissertation. Additional information can be found in the in the following three publications, which were the foundation for my dissertation.

3.1 Animals

In this project, liver fibrosis was induced in male 7-21 weeks old BALB/cANCrl mice by BDL or CCl_4 injection. Mice of this strain were purchased from Charles River (Wilmington, MA USA) and bred in the facilities of Rostock University Medical Center under conditions free of specific pathogens. Mice were given more than 2 days of acclimatization time before the start of the experiment. They were kept in groups before the actual experiment, but were single housed in a European standard type III plastic cage during the experiment at 21 ± 2°C , 12h / 12h of light and dark cycle (light time: 07.00h-19.00h), 60 ± 20% relative humidity, with food (pellets, V1534.000, 10 mm, ssniff) and water provided ad libitum. Enrichment was provided with shredded tissue paper (PZN03058052, FSMED Verbandmittel GmbH, Frankenberg, Germany), one paper tunnel (75 x 38 mm, H 0528-151, ssniff) and a wooden enrichment tool (Espe size S, 40 x 16 x10 mm, H0234.NSG, Abedd, Vienna, Austria). All animal experiments complied with the requirements of the European Directive 2010/63 / EU and the laws of Germany, and have been approved by local ethics committees and public authority. (Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei Mecklenburg-Vorpommern)

3.2 Surgical induction of liver fibrosis

In order to induce liver fibrosis by BDL, the mice were quickly anesthetized by inhaling 5

vol. % isoflurane (CP-pharma, Burgdorf, Germany). The mice were then put on a warming plate at 37°C and inserted in a mask supplying them with 1.5 vol. % isoflurane. The mice were then checked, if they were under deep anesthesia (regular spontaneous breathing, no reflex and response after pain stimuli between toes).[46] The eyes were kept wet by eye ointment. 5 mg/kg Carprofen (Rimadyl®, Pfizer GmbH, Berlin, Germany) was injected subcutaneously 5-15 minutes before operation for perioperative analgesia. The abdominal skin was shaven and the legs were fixed. A laparotomy with length of approximately 2 cm in the mid-line of the abdomen was performed. The common bile duct was exposed and carefully separated from the portal vein and the hepatic artery. Then the bile duct was ligated 3 times with 5-0 silk and cut between the two distal knots.[47] Both abdominal layers (peritoneum and skin) were closed separately by 6-0 and 4-0 prolene sutures. The mice were placed in a cage warmed by a heating lamp until completely awake and active. After the operation, wet pellets were provided as a dietary improvement. Metamizole was added to the drinking water at a concentration of 1250 mg/L, to relieve pain.

3.3 Chemical induction of liver fibrosis

The liver fibrosis model was induced by CCl_4 (Merck Millipore, Eschborn, Germany, code 1.02209.1000). The mice were intraperitoneally injected with a CCl_4/corn oil solution (containing 75% corn oil) at a dosage of 0.25 ml CCl_4/kg body weight twice a week (every Monday and Thursday) for 6 weeks. To reduce pain, 0.25 ml of metamizol (500 mg/mL, Ratiopharm GmbH, Ulm, Germany) was added to the drinking water (100 ml) until the end of the experiment (see Figure 2).

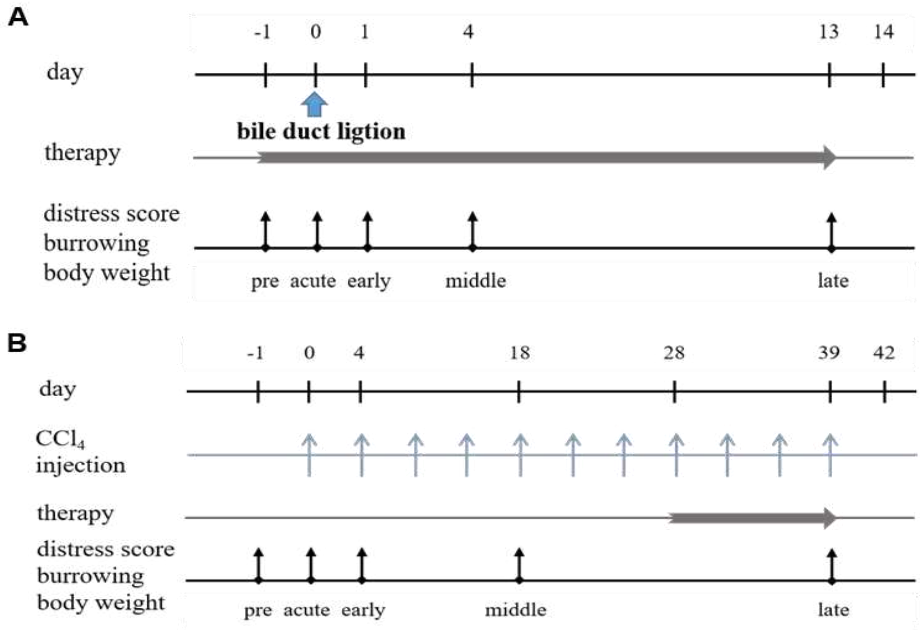


Figure 2. Experimental timelines for distress assessment. Bile duct ligation was performed on day 0

and the distress parameters were evaluated on the indicated days during pre, acute, early, middle and late phase (A). CCl_4 was injected and the distress parameters were evaluated during pre, acute, early, middle and late phase on the indicated days (B).

3.4 Analysis of animal distress and liver damage

3.4.1 Analysis of corticosterone and of liver enzymes in the blood

Blood samples from each mouse were collected 2-4 weeks before any intervention and on the last day before the mice were euthanized. The mice were quickly anesthetized with 5% isoflurane, and 100-150 µl of blood was collected by retro-orbital puncture within 3 minutes to avoid the increase of corticosterone caused by handling and anaesthesia. Blood samples were collected into test tubes, and the plasma was collected by centrifugation and stored at -80°C until measuring the concentration of corticosterone. As parameters for evaluating liver injury, plasma lipase, aspartate aminotransferase (AST) and alanine aminotransferase (ALT) activities were analyzed using a photometer (cobas c111, Roche Diagnostics, Rotkreuz, Switzerland). Analysis was performed by a person, who was unaware of the experimental design and group allocation.

3.4.2 Evaluation of the distress score

The health status of mice was evaluated by multiple clinical parameters as defined in the distress score table (Table 1). When the total score was higher than 15, the mice were euthanized to avoid further deterioration of health and more suffering. The distress score was evaluated regularly and presented before (pre) and after the first intervention (post), during the acute phase (day 0), early phase (BDL: day 1, CCl_4: day 4), middle phase (BDL: day 4, CCl_4: day 18), and late phase (BDL: day 13, CCl_4: day 39) of liver damage. In the CCl_4 model, the distress was always evaluated 30 $\pm$ 5 minutes after the CCl_4 injection or after the end of anaesthesia.

3.4.3 Assessment of natural behaviors

Mutual learning between mice will affect burrowing activity and nesting activity, and social facilitation will enhance this behavior.[48] Therefore, the mice were trained in groups for 3 days and then tested separately.

To assess burrowing activity of mice, a tube (length: 15 cm, diameter: 6.5 cm, closed at one end and raised 3cm at the other) filled with 200 g of food pellets was placed into the left corner of the cage at 16:00, 2-3 hours before the dark phase. After 17 $\pm$ 2 hours, the remaining pellets of the tube were weighed to calculate the amount burrowed. The weight of burrowed pellets was evaluated before the first intervention (pre) and after the first intervention (post) during the acute phase (day 0), early phase (BDL: day 1, CCl_4: day 4),

middle phase (BDL: day 4, CCl_4: day 18), and late phase (BDL: day 13, CCl_4: day 39) of liver damage. The weight of burrowed pellets on day 7 or 8 before intervention was used as a reference for the respective cohort to calculate the change in burrowing activity.

To evaluate nesting activity of mice, the nestlet (5 x 5 cm, cotton) was put into the cage at 16:00-18:00, 2h (+/-1h) before the dark phase. All other forms of nest building material was removed. After 17 ± 2, Nest quality was quantified with a complexity scoring system. Nesting activity was measured before the first intervention (pre) and after the first intervention (post) during the acute phase (day 0), early phase (BDL: day 1, CCl_4: day 4), middle phase (BDL: day 4, CCl_4: day 18) and late phase (BDL: day 13, CCl_4: day 39) of liver damage.

Table 1: Score-Sheet

observations (variables)	score
I Body weight	
I-a decreased > 10% (compared to initial weight)	2
I-b decreased > 20% (compared to initial weight)	5
II General condition	
II-a tooth displacement, too long teeth	1 (A)
II-b fur dull, ruffled or untended	2
II-c eyes unclear or squinted	2
II-d untended orifices of the body	3
II-e abnormal posture	3
II-f dehydration	3
II-g short spasms or temporary paralysis symptoms	3
II-h persistent (>30') cramping or paralysis	5
II-i abnormal respiratory sounds or animal feels cold	5
III Spontaneous behavior	
III-a the animal is passive or overactive	2
III-b pronounced apathy, hyperkinetic, or isolation	4
III-c squeaking due to pain	5
III-d self-mutilation	5
IV Flight behavior after contact	
IV-a animal is passive or overactive	2
IV-b distinct apathy or hyperkinetic	5
V Process-specific criteria	
V-a wound healing disorder	2
V-b opening of the sutures by biting	1 (B)
V-c local inflammation	2
V-d ascites	4
total score	0-66

3.4.4 Body weight analysis

The body weight of mice was evaluated before the first intervention (pre) and after the first intervention (post) during the acute phase (day 1), early phase (BDL: day 2, CCl_4: day 5), middle phase (BDL: day 5, CCl_4: day 19) and late phase (BDL: day 14, CCl_4: day 40) of liver

fibrosis. Body weight was evaluated one day after measuring distress, burrowing or nesting activity. This allows that the mice have enough time to adjust their body weight to a certain stress level.

3.5 Data analysis

For graphing Box plots, line graphs, ROC curves, Kaplan Meier estimators as well as calculating logistic regression and Cox proportional-Hazards model, SigmaPlot 12.0 (SYSTAT Software Inc., San Jose, USA) was used in all studies. Hypothesis test (including Mann-Whitney rank-sum test, Non-parametric test, Shapiro-Wilk normality test, Levene median equal variance test, Student's t-test, Fisher's Exact Test and Log-rank test) were also performed by SigmaPlot 12.0 (SYSTAT Software Inc., San Jose, USA).

In addition, R software with the following packages: caret52 and e107153 was used to build a Support Vector Machine for analyzing distress considering multiple readout parameters simultaneously. R software (Foundation for Statistical Computing, Vienna, Austria) was also used in order to calculate Harrell's concordance indices (C-indices) for determining the optimal prognosis model. Youden's index was used in determining the best cut-off value to distinguish between survivors and non-survivors.[49] For details on data analysis also see Material and Method sections and figure legends in the publications.

4. Results

4.1 Comparing distress based on Corticosterone kinetics

After chemical induction of liver damage by administration of CCl_4, we observed significantly higher corticosterone concentration than in healthy animals after 30, 60, and 120 minutes. The highest concentration was observed after 30 minutes (Figure 3A). To compare stress hormone levels with tissue injury parameters, we measured the time course of plasma AST and ALT activity (Figures 3B and 3C). We observed significant increases in transaminase activity 30, 60, 120, and 240 minutes after CCl_4 injection (Figures 3B and 3C).

After the surgical induction of liver damage by BDL, we observed that significant increases in circulating corticosterone concentrations were detected at 30, 60, 120 and 240 minutes after surgery (Figure 3D). In contrast to CCl_4 injection (Figure 3A), the BDL model induced corticosterone concentration to remain stable at a high level for 120 minutes (Figure 3D). AST and ALT values were already increased significantly at 30min and increased sharply over time (Figures 3E and 3F).

The results suggest that both animal models can induce liver injury and increase corticosterone levels. Compared with the chemically induced animal model (liver damage by CCl_4), the absolute concentration of corticosterone in the BDL model is higher, and the

duration of the corticosterone response is longer. Interestingly, both the absolute value of corticosterone concentration and the duration of the corticosterone response were strongly correlated with the quantification of distress using the score sheet (Table 2).

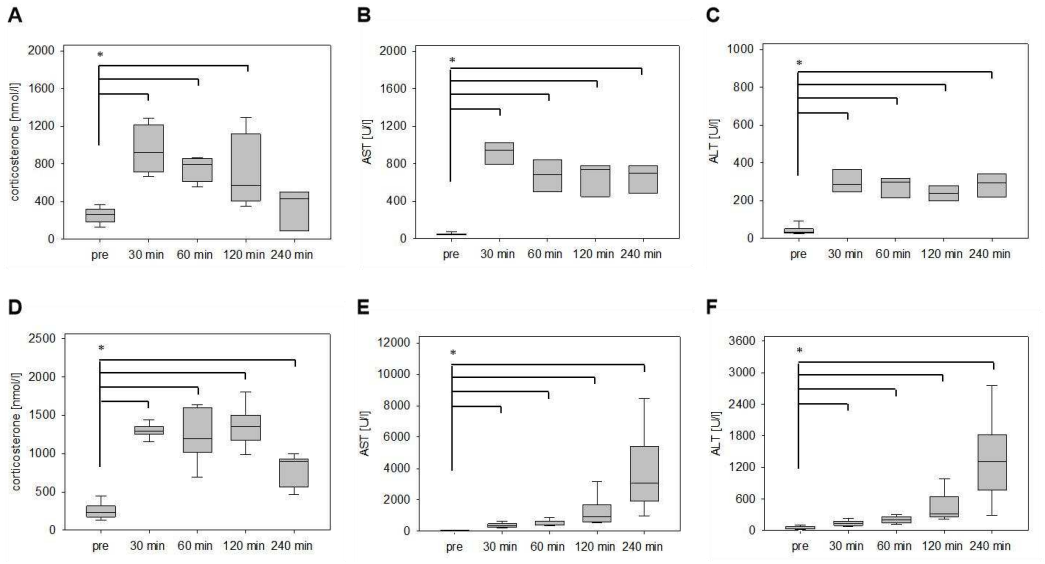


Figure 3. Corticosterone concentration and transaminase activities after chemical induction and surgical induction of liver damage. Analysis of stress hormone release of mice before (pre) and at the indicated time points after the i.p. injection of CCl_4 (0.25 µl/g) (A) as well as the evaluation of AST (B) and ALT(C). Significant differences: *p≤0.0125; n=15 (pre), n=3-4 (30, 60, 120, 240 min). Evaluation of corticosterone concentration of mice before (pre) and after the awakening of the mice from anesthesia (D), as well as the evaluation of transaminase activity of AST (E) and ALT (F) as parameters for the tissue damage. Significant differences: *p≤0.0125; n=20 (pre); n=5 (30, 60, 120, 240 min).

Table 2. Animal models of liver fibrosis ranked according to distress, as well as absolute value and duration of corticosterone concentration.

BALB/c
Distress (mean value of distress scored at 30 min)
BDL > CCl_4
5.0 > 2.0
Corticosterone concentration at 30 min(interquartile range 25-75% in nmol/l)
BDL > CCl_4
1218.1-1383.4 > 710.6-1215.1
Duration of corticosterone response(latest time point in minutes with a significant increase in corticosterone concentration)
BDL > CCl_4
240 > 120

4.2 Comparing distress of animal models

4.2.1 Characterization of methods measuring distress after BDL

We hypothesized that methods are applicable to measure distress, when they can distinguish well between healthy and diseased mice as well as between mice, which

survived (survivors) and mice, which needed to be euthanized (non-survivors). We assessed non-invasive methods such as distress score, burrowing activity and body weight. Data representing healthy mice before (pre-stage) BDL, mice that survived to day 14 after BDL (survivors), and mice that had to be euthanized after BDL (non-survivors) were defined as distress level 0, 1 and 2, respectively.

First, Comparing distress before BDL (pre) with distress after BDL (post), we observed that BDL led to a significant increase in distress scores (Figure 4A), a significant decrease in burrowing activity (Figure 4B), and a decrease in body weight (Figure 4C). ROC curves assessed the performance of these methods in differentiating between distress level 0 and 1. We observed that the distress score, burrowing activity, and body weight could discriminate well between these distress levels (Figure 4D). Combining multiple distress parameters with binary logistic regression, it was found that the combination of these three parameters produced a very high AUC, indicating a very good performance in differentiating between these distress levels (Figure 4E).

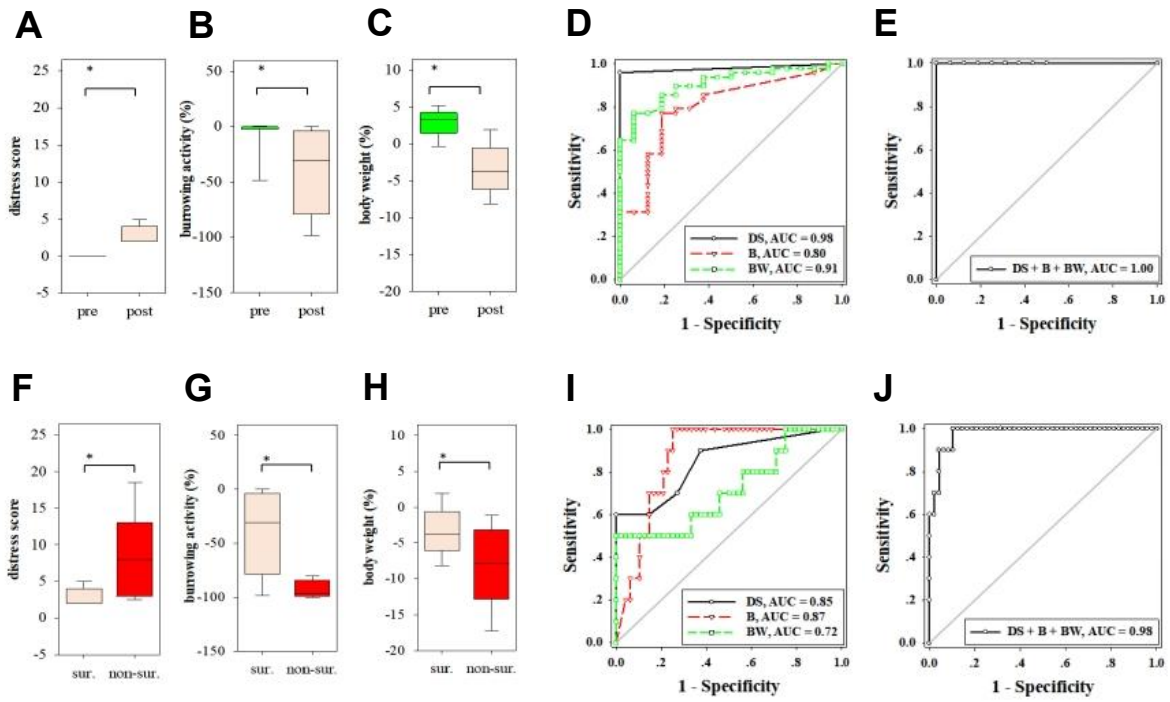
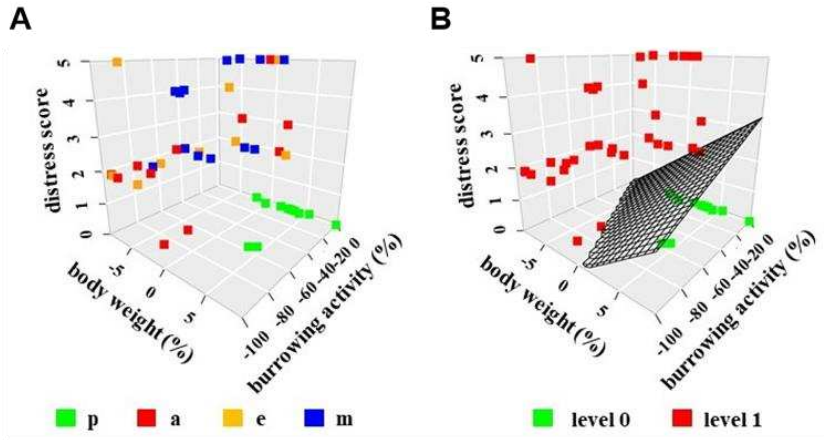


Figure 4. Comparison of distress before (pre) and after (post) BDL (A-E) and between non-survivors just before euthanasia and survivors (F-J). The distress score (A and F), burrowing activity (B and G) and body weight (C and H) was compared between the indicated groups. ROC analysis computed the area under the curve (AUC) for single (D and I), and the combination of all three (E and J) distress parameters (the performance of single and multiple parameters is described by the AUC). 16 mice were used to compare 16 pre data points to 48 post data points (A-E). Distress of 16 Survivors (48 post data points) was compared to distress of 10 non-survivors (10 data points) (F-J). *p< 0.05.

We then explored whether survivors reached a different level of distress than non-survivors. We observed that the non-surviving mice had a significantly increase in the distress score

(Figure 4F), a significantly decrease in burrowing activity (Figure 4G), and a body weight loss (Figure 4H). Using again ROC curves we observed that these methods can discriminate distress level 1 from 2 very well (Figure 4I). Combining these three parameters with binary logistic regression produced a very high AUC, indicating a very good performance in defining distress (Figure 4J).

Next, support vector machine (SVM) was used to classify the samples to distinguish the distress of healthy (pre-intervention) and diseased (post-intervention) mice. Class-labels were obtained by labelling pre-against post-intervention data. For subsequent classification, the data was randomly divided into a training dataset (containing 70% data) and a test dataset (containing 30% data). The training data was used to build the model (Figure 5A). In support vector machines, linear kernel functions are used to find classifiers. This tuned and optimized discriminator was visualized in the plots as a hyperplane that separates the two hypothetical distress levels, defined as distress level 0 or distress level 1 (Figure 5B). Hyper-parameter tuning and 10 cross-validations of 5 replicates were used for internal model optimization and to address potential sampling bias. High average accuracy, sensitivity, and specificity of the model were observed (Figure 5C). This showed that the combination of three parameters (score, burrowing activity, body weight) had a high diagnostic power for the distinction between distress level 0 and pain level 1.



C

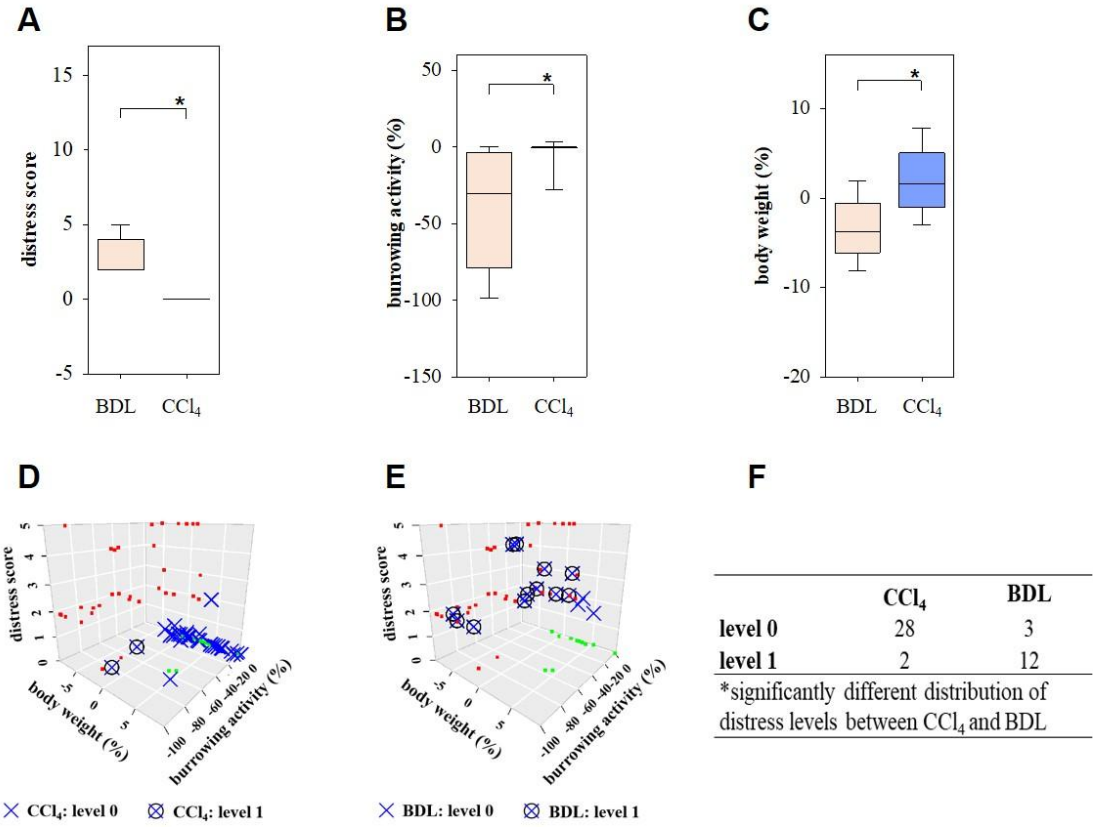
	body weight (%) + burrowing activity (%) + distress score			
	not optimized training data	not optimized test data	optimized training data	optimized test data
accuracy	1	1	1	0.85
sensitivity	1	1	1	1
specificity	1	1	1	0.80

Figure 5. Generation of a training model by SVM. Single data points (squares), which were derived from the training data set from identical animal previous (p) to BDL and at the acute (a), early (e), and middle (m) phase of cholestasis are presented in form of a three dimensional scatter plot (a). A

discriminatory model was built by training a linear SVM kernel to the labelled data in order to differentiate between two levels (level 0 and level 1) of distress (B): The resulting classifier (hyperplane) discriminates between these two levels. The accuracy, sensitivity and specificity of the training model was characterised using either the training data themselves or a test data set and applying the hyperplane (not optimized) or an optimized hyperplane after a 5-times repeated 10-fold cross validation (C). Training data set: n = 11 data points (pre), post: n = 33 data points (post); test data set: n = 5 data points (pre), post: n = 15 data points (post).

4.2.2 Comparing distress of the BDL to the CCl_4 animal model

We compared two animal models that are widely used to study liver injury or fibrosis, the BDL and the CCl_4 animal model. We observed that compared with CCl_4-treated mice, the BDL mice had a significant increase in distress scores (Figure 6A), a significant decrease in burrowing activity (Figure 6B), and a decrease in body weight (Figure 6C).



	CCl_4	BDL
level 0	28	3
level 1	2	12

*significantly different distribution of distress levels between CCl_4 and BDL

Figure 6. Comparing distress caused by CCl_4 injections and bile duct ligation. The distress score (A) was decreased ($P \leq 0.001$), whereas burrowing activity (B) was increased ($P \leq 0.001$) and body weight loss (D) was decreased ($P \leq 0.001$) when comparing post-BDL to post-CCl_4 animals. BDL: 16 mice = 48 data points; CCl4: 10 mice = 30 data points. Using SVM classification, both plots green and red squares indicate distress level 0 or distress level 1 of the BDL training data set and crosses denote data classified as distress level 0, whereas circled crosses denote data classified as distress level 1 (D, E). Blue crosses denote post-CCl_4 (D) and post-BDL (E) distress. A 2x2 contingency table compares the distributions of predicted distress levels of post-CCl_4 to the post-BDL test data set (F). A significantly different distribution of distress levels between these test data sets has been determined by Fisher's Exact Test, *$P \leq 0.001$), CCl_4: n = 30 data points; BDL: n = 15 data points.

Therefore, all three read-out parameters indicate that BDL causes more distress than CCl_4.

Then we compared the two animal models using an optimized training model based on 70% of the BDL data. We classified the data after CCl_4 and after BDL according to this training model (Figures 6D and 6E). We observed that only 2 out of 30 post-CCl_4 data points were assigned to distress level 1, while 12 out of 15 post-BDL data points were correctly assigned to distress level 1 (Figure 6F). Using Fisher's exact test, a significant difference in the distribution of was observed between the BDL and CCl_4 cohorts ($P < 0.001$). This multivariate analysis showed that at most time points, animals treated with CCl_4 experienced less distress than animals after a BDL.

4.3 Improving euthanasia of mice after bile duct ligation

4.3.1 Retrospective analysis of distress parameters

In order to define early humane endpoints, score sheet criteria (II-a, II-d, II-g, II-h, III-c, III-d, V-a, V-b, V-c), which were never observed after bile duct ligation and humane endpoint criteria (I-b, II-i, IV-b), which lead to the immediate euthanasia of the animals, were excluded from analysis (Figure 7).

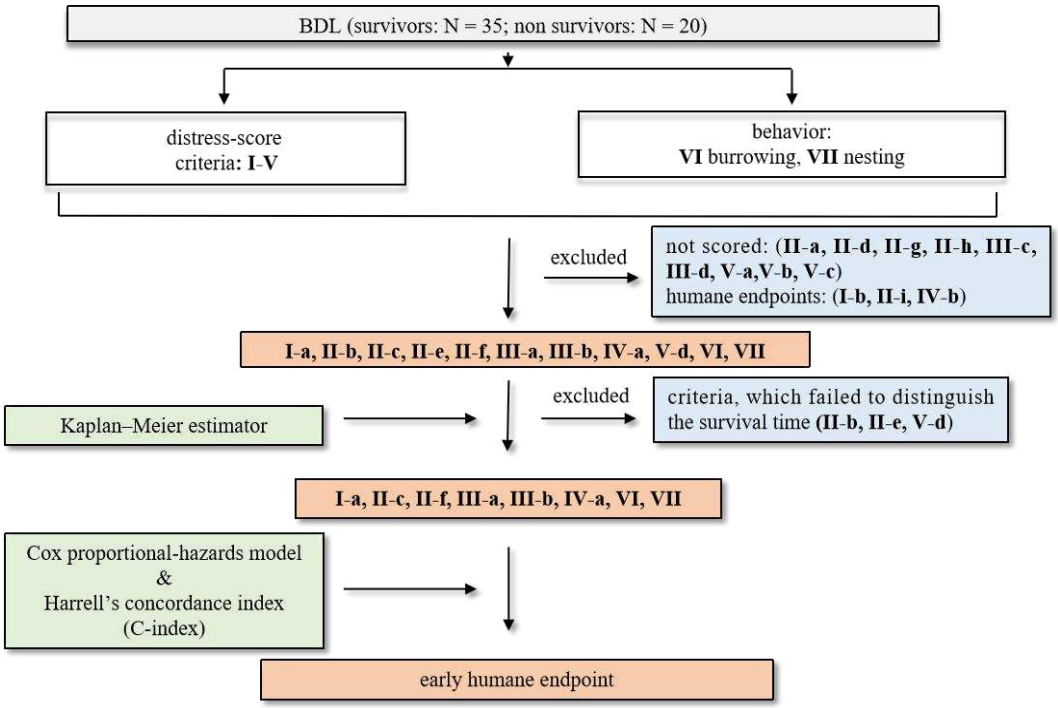


Figure 7. Flowchart to retrospectively analyse distress parameters in order to determine an early humane endpoint. First, score sheet criteria (Table 1), which were not observed during the experiment as well as humane endpoint criteria, were excluded. Second, Kaplan-Meier estimator curves were used to exclude criteria, which did not predict survival time. Third, the performance of each single as well as combinations of multiple parameters were analyzed by Cox proportional-hazards model followed by Harrell's concordance index, to determine which criteria combination might be used as efficient early humane endpoint.

Then, eleven variables were evaluated by the Kaplan-Meier estimator (Figure 8). Criteria (II-b, II-e, V-d) with no significant difference in survival time were also excluded from further analysis (Figures 8 A-C). Mice with a positive status of specific criteria (I-a, II-c, II-f, III-a, III-b, IV-a, VI or VII) had a significantly ($P \leqslant 0.005$) shorter survival time than mice, where these criteria were not observed (Figure 8 D-K).

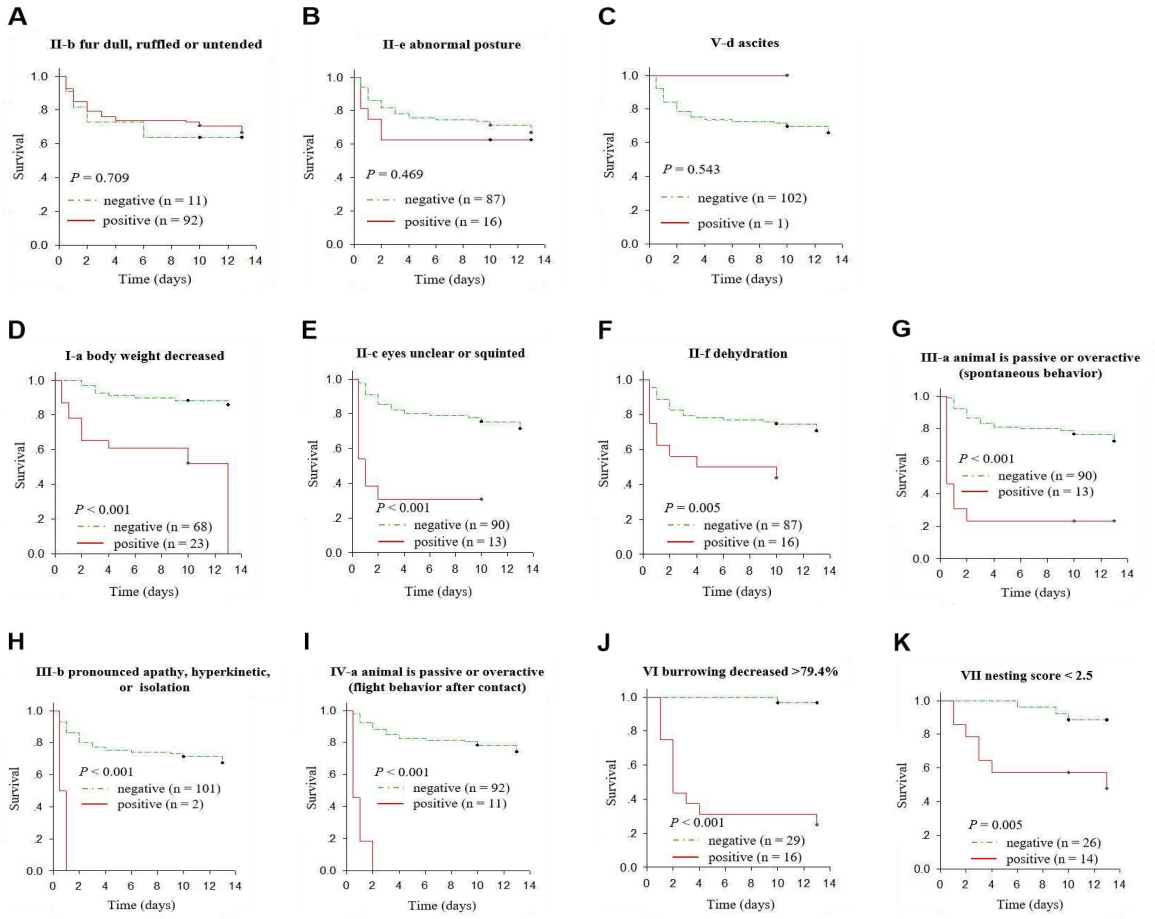


Figure 8. Kaplan-Meier curves using distinct score sheet criteria and read-out parameters for distress. The positive status of II-b (fur dull, ruffled or untended) (A), II-e (abnormal posture) (B) or V-d (ascites) (C) failed to significantly indicate reduced survival time of mice. The P-value was determined by log-rank test. The positive status of I-a (body weight decreased 10 to 20 %) (D), II-c (eyes unclear and squinted) (E), II-f (dehydration) (F), III-a (spontaneous behavior: animal is passive and over active) (G), III-b (pronounced apathy, hyperkinetic or isolation) (H), IV-a (flight behavior after contact: animal is passive or overactive) (I), decreased burrowing activity by > 79.4 % (J) and nesting score of less than 2.5 (K), a significantly shorter survival time was calculated by log rank test ($P \leq 0.005$).

4.3.2 Predictive models for defining survival time

The univariate Cox proportional-hazards model was used to assess the hazard ratio (HR) of each variable (Table 3). All tested criteria (I-a, II-c, II-f, III-a, III-b, IV-a, VI or VII) significantly increased the risk of death. As a next step, a multivariate analysis was used to assess the hazard ratio (HR) of these distress score criteria (strategy 1 in Table 3). The positive status

of two criteria, I-a and IV-a, significantly increased the risk of death. Then, two behavioral parameters (burrowing and nesting activity) were included in our analysis. These two variables, assessed in two different cohorts, were analyzed in two additional multivariate Cox proportional risk models (Strategy 2 and Strategy 3). First, we added burrowing and nesting activity to all distress score variables, respectively (strategy 2 and strategy 3). Positive status of variables I-a (body weight) and VI (burrowing activity) was observed to significantly increase the risk of death (strategy 2 in Table 3). In another multivariate Cox proportional-hazards model that uncluded nesting activity also a significantly increased risk of death was noticed (strategy 3 in Table 3).

Table. 3 HR (hazard ratio) and P-value (P) for the distinct variables were determined by Cox proportional-hazards model.

variables		univariate		multivariate					
		HR	*P*	strategy 1		strategy 2		strategy 3	
				HR	*P*	HR	*P*	HR	*P*
I-a	negative	reference		reference		reference		NS	
	positive	6.169	< 0.001	3.968	0.017	24.096	0.007		
II-c	negative	reference		NS		NS		-	
	positive	5.033	< 0.001						
II-f	negative	reference		NS		NS		NS	
	positive	2.825	0.009						
III-a	negative	reference		NS		NS		NS	
	positive	6.584	< 0.001						
III-b	negative	reference		NS		-		-	
	positive	9.495	0.003						
IV-a	negative	reference		reference		NS		-	
	positive	17.130	< 0.001	8.621	0.021				
VI	negative	reference		-		reference		-	
	positive	33.218	< 0.001			54.348	0.003		
VII	negative	reference		-		-		reference	
	positive	5.639	0.013					5.051	0.026

NS indicates no significant difference (P > 0.05).

4.3.3 Evaluation of models for defining survival time

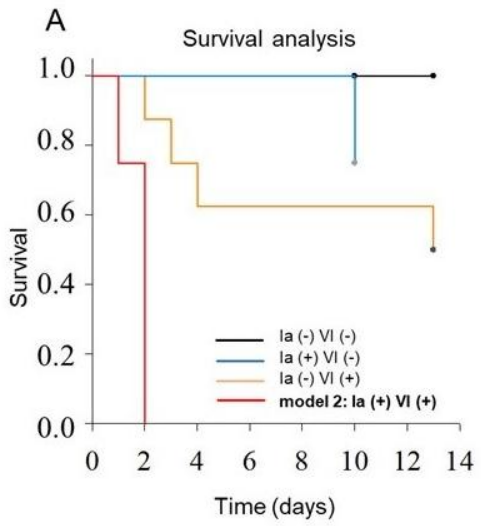
Harrell's Concordance index was then used to compare models of single variables or combinations of variables. The indices of each model was bootstrapped to get more robust estimates, thereby internally validating the goodness-of-fit (Table 4) of each model. We observed that model 2 (C-index: 0.947, 95 % CI: 0.832-0.978) has a higher C-index than model 1 (C-index: 0.720, 95 % CI: 0.590-0.838) or model 3 (C-index: 0.724, 95 % CI: 0.532-0.849). In addition, the model 4 which combined all distress score parameters was evaluated

in the univariate Cox proportional-hazards model. We observed that model 2 also had a higher C-index than model 4 (C-index: 0.726, 95 % CI: 0.553-0.782) (Table 4). The C-index of Model 2, body weight plus burrowing activity (C-index: 0.947, 95% CI: 0.832 - 0.978) was also higher than that of all single variables (Table 3). The C-index ranges from 0.5 to 1, where 0.5 is completely random, indicating that the model has no predictive effect, and 1 is completely consistent, indicating that the predicted results of the model are completely consistent with the reality. Since the C-index as well as the bootstrapped C-index of our model 2 is above 0.9, this suggests that model 2 can very well predict the survival times of these animals.

Table 4 Bootstrapped C-indices (Harrell's concordance index) and the corresponding 95% confidence intervals for each distinct model.

	variables	C-index	C-index bootstrapped	95% CI
single variables	I-a	0.691	0.692	0.572-0.797
	II-c	0.633	0.632	0.553-0.721
	II-f	0.590	0.590	0.517-0.680
	III-a	0.657	0.656	0.574-0.746
	III-b	0.534	0.534	0.500-0.614
	IV-a	0.684	0.683	0.603-0.770
	VI	0.865	0.866	0.768-0.923
	VII	0.725	0.724	0.534-0.848
model-1	Ia, IVa	0.719	0.720	0.590-0.838
model-2	**Ia, VI**	**0.943**	**0.947**	**0.832-0.978**
model-3	VII	0.725	0.724	0.532-0.849
model-4	I-a, II-c, II-f, III-a, III-b, IV-a	0.696	0.726	0.553-0.782

Kaplan-Meier estimates were used to evaluate the practicality of the combination. We observed that mice, which lost body weight in a range of 10 % to 20 % and also showed decreased burrowing activity by more than 79.4 % had a significantly shorter survival time comparing to mice which were positive for only one or none of these two variables (Figure 9). All mice with such a reduction in body weight and burrowing activity died within 2 days (Figure 9).



B

P-values	Ia (-) VI(-)	Ia (+) VI(-)	Ia (-) VI(+)
Ia (-) VI(-)	-	-	-
Ia (+) VI(-)	0.025	-	NS
Ia (-) VI(+)	0.002	NS	-
Ia (+) VI(+)	< 0.001	0.030	0.014

Figure 9 Kaplan-Meier curve of the multivariate Model 2: Mice with a body weight loss of more than 10 to 20 % (variable I-a) and decreased burrowing activity of more than 79.4 % (variable VI) died within 2 days (A). Mice positive for variable Ia and VI had a significantly decreased survival time when compared to mice positive for just one variable or negative for both variables (B). The P-values were determined by log-rank test ($P < 0.05$).

5. Discussion

5.1 Evaluating Corticosterone kinetics for defining animal distress

The results of this study suggest that the absolute value of corticosterone concentration and the duration of corticosterone response are suitable parameters to compare the severity in different animal models. When comparing the absolute value of corticosterone concentration and the duration of corticosterone response between the two animal models, a very similar rank order could be observed (Table 2). In addition this study suggests that the absolute value of corticosterone concentration and the duration of corticosterone response are closely related to the quantification of animal distress by a score sheet. Some studies support the hypothesis of a direct correlation between stressor intensity and corticosterone concentration.[50-52] Therefore, we recommend that, the two parameters, absolute value of corticosterone concentration and duration of the corticosterone response, should be regarded as appropriate indicators for quantifying distress.

Acute liver damage was induced by chemical induction (CCl_4 intraperitoneal injection) and surgical intervention (BDL) in BALB/c mice (Figure 3 B-C and E-F). Both interventions caused a significant increase in corticosterone concentrations within 30 minutes. However, compared with the transient increase in corticosterone concentration after CCl_4 injection, the concentration of corticosterone increased significantly for a longer period after BDL (see Figures 3 A and D). It has also been reported in some studies that a long-lasting corticosterone responses may be characteristic of post-operative stress responses.[50, 53, 54] The causes of a long lasting corticosterone responses may include anesthesia, pain, wound

tension, tissue damage, and release of cytokines. Some studies reported that, a significant increase in corticosterone concentration within minutes of anesthesia using isoflurane. However, the increased corticosterone concentration was transient and declined already after 60 min. Thus, anesthesia alone is probably not responsible for the long lasting activation of the HPA axis. Cytokine IL-6, which can activate the HPA axis to release corticosterone, usually was released within 60 minutes after surgical invention.[55-58] Moreover, it was positively correlated with corticosterone concentration.[58, 59] Therefore, we suggest that the long lasting corticosterone response after surgery might be caused by several independent mechanisms that influence the HPA axis.

Our study demonstrates that corticosterone kinetics can be used as an important parameter for comparing distress levels between different animal models. However, corticosterone is also affected by other physiological responses, such as the circadian rhythm,[15, 60, 61] the estrus cycle,[15, 62, 63] or sexual arousal.[64] Therefore, it is recommended that corticosterone should be evaluated together with other stress parameters, such as assessing a distress score or body weight. Different stressors reach maximum corticosterone concentrations at different time points, e.g., 20-30 minutes after injection of toxic chemicals,[50, 65] whereas maximum corticosterone concentrations are observed 60 minutes after electric shock[66] or acute shaker stress.[67] Therefore, prior to choosing a time point for blood analysis, it is necessary to understand the corticosterone kinetics in an animal model. Our data suggest that assessing the corticosterone concentration at 30 minutes is suitable for comparing distress in models of acute liver damage.

5.2 Comparing distress of different animal models

Study II compared distress of mouse models for BDL and CCl_4-induced liver damage. Based on three different non-invasive methods (distress score, body weight and burrowing activity), a multivariate analysis demonstrated that BDL caused more distress than treatment with CCl_4. Methods with high discriminatory power are a prerequisite to evaluate animal distress and to differentiate between distinct distress levels. In this study, ROC curve analysis was used to evaluate the discriminatory power of parameters. ROC curves have been widely used in clinical research to analyze and evaluate diagnostic test data.[68-70] For example, ROC curve analysis contributes to the analysis and evaluation of biomarkers in predicting pancreatic cancer.[71]

In our study, ROC curves were used to analyze and evaluate the suitability of parameters to distinguish between healthy and diseased mice, or diseased mice that survived and diseased mice that would die from the disease. All parameters: distress score, burrowing activity, and weight change had a high discriminatory power (Figure 4D and 4I). When differentiating between health and disease mice, burrowing activity was the parameter with

the lowest performance (performance of parameters: distress score > body weight change > burrowing activity). When differentiating between survivors and non-survivors, but body weight change was the parameter with the lowest performance (performance of parameters: burrowing activity > distress score > body weight change). This suggests that all three parameters are useful for evaluating distress. However, this also demonstrates that not a single parameter can be regarded as “gold standard” for all situations.

In our study, we plotted three parameters and defined distress levels through support vector machine classification. This classification method, considering all three parameters and had a high performance, characterized by a very high AUC (Figures 4E and 4J).

The study relied on physical parameter evaluation, distress score, and body weight and natural behavior, such as burrowing activity. Many studies have shown that body weight is very useful for evaluating animal’s suffering.[23-25] The advantage of body weight is that it is suitable for the assessment of well-being in many species and it is easy to measure. However, the adaption of body weight after an intervention or during a disease takes approximately 24h, this criteria is not practical for acute and severe diseases. It is well known that burrowing behavior is a sensitive indicator of animal suffering in models of surgical intervention, chronic diseases, or neurological disorders.[21, 72, 73] One disadvantage of using burrowing activity as method to measure distress is that it is only applicable for rodents. While our results suggest that support vector machines can be used to compare distress, measured by several methods, in two animal models, it is too early to decide, if this method is appropriate to determine severity in all animal models. This approach still needs to be tested by more research institutions and applied in many different animal models, before reaching a final conclusion.

5.3 Improving euthanasia in a cholestasis model

Our study revealed that it was feasible to use retrospective analysis to analyze the score sheet and additional distress parameters in order to determine an early humane endpoint. Mice, which lost 10 - 20% of body weight and reduced burrowing activity by more than 79.4% died within 2 days. Thus, these cut-off criteria may reduce the suffering of laboratory animals in subsequent experiments and might contribute to the refinement of animal research.

The approach is based on three steps (Figure 7). First, score sheet criteria, which are not observed in the experiment are excluded. Second, criteria, which do not contribute to the prediction of survival time are excluded (Kaplan-Meier Method). Third, the performance of single parameter and multiple parameter combination is analyzed (Cox proportional-hazards model and Harrell’s concordance index). Interestingly, survival analysis using Kaplan-Meier estimators showed that mice with the performance of III-b (marked indifference, hyperactivity,

or isolation) or IV-a (animal passivity or hyperactivity-flight behavior after contact) died within 1-2 days (Figures 8 H and I). Therefore, these two indicators could also be considered as criteria for determining early humanitarian endpoints. However, the C-index (III-b: 0.534, IV-a: 0.684) of these two variables is quite low compared to multivariate model 2 (C-index: 0.943), and the incidence of III-b (n = 2) is also very low. This indicates that multivariate distress analysis is superior to univariate analysis. This conclusion is consistent with the results of other studies using multiple parameters to predict death.[24, 25, 74]

The Cox proportional-hazard model and Harrell's concordance index of models are commonly used in statistical medical research.[75-77] It is mainly used to analyze the survival time of patients and single or multiple influencing factors. Using this approach, different univariate and multivariate models can be defined. A comparison of the C-indices revealed that the combination of body weight loss and reduction of burrowing behavior is the best model for predicting survival times (Table 3). Model 2 has a C-index well above 0.9. C-index ranges from 0.5 to 1, where C = 0.5 represents complete randomness, while C = 1 reveals perfect prediction rules.[23] Since both the C-index and the bootstrapped C-index of our model 2 is above 0.9, we conclude that model 2 can well predict the survival times of murine animals.

The two variables, body weight and burrowing activity, can be measured objectively, which minimize potential selection bias. And the two variables are also considered reliable indicators of measuring suffering of mice, as reported in many studies.[35-37, 78, 79] However, the concept that a combination of these two variables can define a good humane endpoint should be externally validated in future studies, before it is widely applied.

For different animal models, several univariate and multivariate analysis studies have been reported to determine early humane endpoints.[80] For example, it was suggested that a combination of body weight loss and reduced body temperature could serve as early humane endpoint in a mouse model of ocular herpes-virus Infection.[25] It is worth mentioning that most studies so far do not include time as a factor or provide information on how quickly the animals died.[80] Thus, compared to the already published methods for early humane endpoint determination the advantage of our approach is the inclusion of the time variable by Cox regression and that a justifiable low number of animals is required for such an analysis.

5.4 Comparing the pathogenic mechanism between BDL and CCl_4 mouse models

Although liver fibrosis can be induced by various approaches, such as BDL or administration of CCl_4, some aspects of fibrogenesis are similar. Following hepatocyte damage, various stimuli, such as oxidative stress or cytokines cause the transdifferentiate of resting hepatic stellate cell into liver myofibroblasts. This is mediated directly or indirectly by immune cells such as Kupffer cell and monocytes and results in the deposition of extracellular matrix and

destruction of the architecture of the liver (see Figure 1).[44, 81, 82] Transforming growth factor b1 (TGF-b1), as an important cytokine in liver fibrogenesis, can activiate resting hepatic stellate cells and regulate extracellular matrix deposition.[44, 83] Interestingly, a TGF-b1 inhibitor has been reported to significantly reduce the amount of fibrosis in BDL rats.[84] This indicate that TGF-b1 play an important role in this cholestatic fibrosis model. Like in the model of cholestatic fibrosis, blocking TGF-b1 also significantly reduces fibrosis in CCl_4-induced liver fibrosis.[85] This indicates that the BDL and the CCl_4 models induce fibrosis at least partially via similar mechanisms. However, there are also differences between these two animal models, In the BDL model, the toxicity of bile salts can induce hepatocyte apoptosis, and leads to an oxidative stress response via activation of NADPH oxidase isoforms.[43, 86] In addition, complications, such as intraoperative or postoperative bleeding and severe infections, can occur during and after bile duct ligation.[46] In the CCl_4 model, after repetitively administration of CCl_4, CCl_3 radicals are formed, which induce hepatocyte damage and activation of hepatic stellate cells. In contrast to bile salts, CCL_4 is not only a strong liver toxin, but is also a powerful nephrotoxic and carcinogenic agent.[87] When choosing an animal model for preclinical research, the similarity of pathophysiological features between the animal model and the humane disease but also the distress caused to the animals should be considered.

6. Conclusion and outlook

This dissertation demonstrates that some methods, such as assessing the peak and duration of corticosterone concentration or evaluating a distress score sheet, burrowing activity, and body weight change, are informative, when evaluating distress in liver fibrosis models. Using these sensitive indicators, we evaluated and compared distress in distinct animal models, as well as identified early humane endpoints in cholestatic mice. I hope that my work will draw the attention of clinicians and scientists to animal welfare so that they will integrate the assessment of animal suffering in their preclinical research. This will reduce animal suffering in future experiments and develop improved animal models.

7. Reference

1. Thèves, C.; Biagini, P.; Crubézy, E., The rediscovery of smallpox. *Clin Microbiol Infect* **2014,** *20* (3), 210-8.
2. Clemens, J. D.; Nair, G. B.; Ahmed, T.; Qadri, F.; Holmgren, J., Cholera. *Lancet* **2017,** *390* (10101), 1539-1549.
3. Pucchio, A. M. R.; Alabdulraheem, A.; Salvadori, M. I., Polio. *Cmaj* **2022,** *194* (44), E1509.
4. Timsit, M. O.; Kleinclauss, F.; Thuret, R., [History of kidney transplantation surgery]. *Prog Urol* **2016,** *26* (15), 874-881.
5. Hegele, R. A.; Maltman, G. M., Insulin's centenary: the birth of an idea. *Lancet Diabetes Endocrinol* **2020,** *8* (12), 971-977.
6. Foundation of biomedical research, nobel prizes in physiology or medicine, URL: https://fbresearch.org/medical-advances/nobel-prizes, **accessed on 9th of September 2023**.
7. EU - The European Parliament and the Council of the European Union. Directive 2010/63/EU of the European Parliament and of the Council of 22 of September 2010 on the protection of animals used for scientific purposes. **2010/63/EU**. Offical Journal of the European Union (L-276), 33–79.
8. Bundesministerium der Justiz, Germany （ 2013, Tierschutzgesetz, TierSchG, vom 04.07.2013 (BGBl.S1950). In: BGBl. http://www.gesetze-im-internet.de/tierschg/BJNR012770972.html, **accessed on September, 2023**.
9. Bayne, K.; Würbel, H., The impact of environmental enrichment on the outcome variability and scientific validity of laboratory animal studies. *Rev Sci Tech* **2014,** *33* (1), 273-80.
10. Carbone, L.; Austin, J., Pain and Laboratory Animals: Publication Practices for Better Data Reproducibility and Better Animal Welfare. *PLoS One* **2016,** *11* (5), e0155001.
11. Leenaars, C. H. C.; van der Mierden, S.; Durst, M.; Goerlich-Jansson, V. C.; Ripoli, F. L.; Keubler, L. M.; Talbot, S. R.; Boyle, E.; Habedank, A.; Jirkof, P.; Lewejohann, L.; Gass, P.; Tolba, R.; Bleich, A., Measurement of corticosterone in mice: a protocol for a mapping review. *Lab Anim* **2020,** *54* (1), 26-32.
12. Gjendal, K.; Ottesen, J. L.; Olsson, I. A. S.; Sørensen, D. B., Burrowing and nest building activity in mice after exposure to grid floor, isoflurane or ip injections. *Physiol Behav* **2019,** *206*, 59-66.
13. Roughan, J. V.; Flecknell, P. A., Evaluation of a short duration behaviour-based post-operative pain scoring system in rats. *Eur J Pain* **2003,** *7* (5), 397-406.
14. Graf, R.; Cinelli, P.; Arras, M., Morbidity scoring after abdominal surgery. *Lab Anim* **2016,** *50* (6), 453-458.
15. Gong, S.; Miao, Y. L.; Jiao, G. Z.; Sun, M. J.; Li, H.; Lin, J.; Luo, M. J.; Tan, J. H., Dynamics and correlation of serum cortisol and corticosterone under different physiological or stressful conditions in mice. *PLoS One* **2015,** *10* (2), e0117503.
16. Munck, A.; Guyre, P. M.; Holbrook, N. J., Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. *Endocr Rev* **1984,** *5* (1), 25-44.
17. Benedetti, M.; Merino, R.; Kusuda, R.; Ravanelli, M. I.; Cadetti, F.; dos Santos, P.; Zanon, S.; Lucas, G., Plasma corticosterone levels in mouse models of pain. *Eur J Pain* **2012,** *16* (6), 803-15.
18. 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* **1985,** *116* (16), 431-6.
19. Smith, D.; Anderson, D.; Degryse, A. D.; Bol, C.; Criado, A.; Ferrara, A.; Franco, N. H.; Gyertyan, I.; Orellana, J. M.; Ostergaard, G.; Varga, O.; Voipio, H. M., Classification and reporting of severity experienced by animals used in scientific procedures: FELASA/ECLAM/ESLAV Working Group report. *Lab Anim* **2018,** *52* (1_suppl), 5-57.

20. Hawkins, P.; Morton, D. B.; Burman, O.; Dennison, N.; Honess, P.; Jennings, M.; Lane, S.; Middleton, V.; Roughan, J. V.; Wells, S.; Westwood, K., A guide to defining and implementing protocols for the welfare assessment of laboratory animals: eleventh report of the BVAAWF/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Lab Anim* **2011,** *45* (1), 1-13.
21. Deacon, R. M.; Raley, J. M.; Perry, V. H.; Rawlins, J. N., Burrowing into prion disease. *Neuroreport* **2001,** *12* (9), 2053-7.
22. Jirkof, P., Burrowing and nest building behavior as indicators of well-being in mice. *J Neurosci Methods* **2014,** *234*, 139-46.
23. Takayama-Ito, M.; Lim, C. K.; Nakamichi, K.; Kakiuchi, S.; Horiya, M.; Posadas-Herrera, G.; Kurane, I.; Saijo, M., Reduction of animal suffering in rabies vaccine potency testing by introduction of humane endpoints. *Biologicals* **2017,** *46*, 38-45.
24. Trammell, R. A.; Toth, L. A., Markers for predicting death as an outcome for mice used in infectious disease research. *Comp Med* **2011,** *61* (6), 492-8.
25. Hankenson, F. C.; Ruskoski, N.; van Saun, M.; Ying, G. S.; Oh, J.; Fraser, N. W., Weight loss and reduced body temperature determine humane endpoints in a mouse model of ocular herpesvirus infection. *J Am Assoc Lab Anim Sci* **2013,** *52* (3), 277-85.
26. Van der Mierden, S.; Leenaars, C. H. C.; Boyle, E. C.; Ripoli, F. L.; Gass, P.; Durst, M.; Goerlich-Jansson, V. C.; Jirkof, P.; Keubler, L. M.; Talbot, S. R.; Habedank, A.; Lewejohann, L.; Tolba, R. H.; Bleich, A., Measuring endogenous corticosterone in laboratory mice - a mapping review, meta-analysis, and open source database. *Altex* **2021,** *38* (1), 111-122.
27. Ingle, D. J., Permissibility of hormone action; a review. *Acta Endocrinol (Copenh)* **1954,** *17* (1-4), 172-86.
28. Morton, D. B., A systematic approach for establishing humane endpoints. *Ilar j* **2000,** *41* (2), 80-6.
29. Nørgaard, S. A.; Sand, F. W.; Sørensen, D. B.; Abelson, K. S.; Søndergaard, H., Softened food reduces weight loss in the streptozotocin-induced male mouse model of diabetic nephropathy. *Lab Anim* **2018,** *52* (4), 373-383.
30. Adamson, T. W.; Diaz-Arevalo, D.; Gonzalez, T. M.; Liu, X.; Kalkum, M., Hypothermic endpoint for an intranasal invasive pulmonary aspergillosis mouse model. *Comp Med* **2013,** *63* (6), 477-81.
31. Gavin, H. E.; Satchell, K. J. F., Surface hypothermia predicts murine mortality in the intragastric Vibrio vulnificus infection model. *BMC Microbiol* **2017,** *17* (1), 136.
32. Deacon, R. M., Assessing nest building in mice. *Nat Protoc* **2006,** *1* (3), 1117-9.
33. Deacon, R. M., Burrowing in rodents: a sensitive method for detecting behavioral dysfunction. *Nat Protoc* **2006,** *1* (1), 118-21.
34. Jirkof, P.; Fleischmann, T.; Cesarovic, N.; Rettich, A.; Vogel, J.; Arras, M., Assessment of postsurgical distress and pain in laboratory mice by nest complexity scoring. *Lab Anim* **2013,** *47* (3), 153-61.
35. Jirkof, P.; Leucht, K.; Cesarovic, N.; Caj, M.; Nicholls, F.; Rogler, G.; Arras, M.; Hausmann, M., Burrowing is a sensitive behavioural assay for monitoring general wellbeing during dextran sulfate sodium colitis in laboratory mice. *Lab Anim* **2013,** *47* (4), 274-83.
36. Deacon, R. M.; Reisel, D.; Perry, V. H.; Nicholas, J.; Rawlins, P., Hippocampal scrapie infection impairs operant DRL performance in mice. *Behav Brain Res* **2005,** *157* (1), 99-105.
37. Pfeiffenberger, U.; Yau, T.; Fink, D.; Tichy, A.; Palme, R.; Egerbacher, M.; Rülicke, T., Assessment and refinement of intra-bone marrow transplantation in mice. *Lab Anim* **2015,** *49* (2), 121-31.
38. Sliepen, S. H. J.; Diaz-Delcastillo, M.; Korioth, J.; Olsen, R. B.; Appel, C. K.; Christoph, T.; Heegaard, A. M.; Rutten, K., Cancer-induced Bone Pain Impairs Burrowing Behaviour in Mouse and Rat. *In Vivo* **2019,** *33* (4), 1125-1132.
39. Campana, L.; Iredale, J. P., Regression of Liver Fibrosis. *Semin Liver Dis* **2017,** *37* (1), 1-10.

40. Urban, M. L.; Manenti, L.; Vaglio, A., Fibrosis--A Common Pathway to Organ Injury and Failure. *N Engl J Med* **2015,** *373* (1), 95-6.
41. Mokdad, A. A.; Lopez, A. D.; Shahraz, S.; Lozano, R.; Mokdad, A. H.; Stanaway, J.; Murray, C. J.; Naghavi, M., Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. *BMC Med* **2014,** *12*, 145.
42. Asrani, S. K.; Devarbhavi, H.; Eaton, J.; Kamath, P. S., Burden of liver diseases in the world. *J Hepatol* **2019,** *70* (1), 151-171.
43. Yanguas, S. C.; Cogliati, B.; Willebrords, J.; Maes, M.; Colle, I.; van den Bossche, B.; de Oliveira, C.; Andraus, W.; Alves, V. A. F.; Leclercq, I.; Vinken, M., Experimental models of liver fibrosis. *Arch Toxicol* **2016,** *90* (5), 1025-1048.
44. Ravichandra, A.; Schwabe, R. F., Mouse Models of Liver Fibrosis. *Methods Mol Biol* **2021,** *2299*, 339-356.
45. Liedtke, C.; Luedde, T.; Sauerbruch, T.; Scholten, D.; Streetz, K.; Tacke, F.; Tolba, R.; Trautwein, C.; Trebicka, J.; Weiskirchen, R., Experimental liver fibrosis research: update on animal models, legal issues and translational aspects. *Fibrogenesis Tissue Repair* **2013,** *6* (1), 19.
46. Tag, C. G.; Sauer-Lehnen, S.; Weiskirchen, S.; Borkham-Kamphorst, E.; Tolba, R. H.; Tacke, F.; Weiskirchen, R., Bile duct ligation in mice: induction of inflammatory liver injury and fibrosis by obstructive cholestasis. *J Vis Exp* **2015,** (96).
47. Abshagen, K.; König, M.; Hoppe, A.; Müller, I.; Ebert, M.; Weng, H.; Holzhütter, H. G.; Zanger, U. M.; Bode, J.; Vollmar, B.; Thomas, M.; Dooley, S., Pathobiochemical signatures of cholestatic liver disease in bile duct ligated mice. *BMC Syst Biol* **2015,** *9*, 83.
48. Deacon, R., Assessing burrowing, nest construction, and hoarding in mice. *J Vis Exp* **2012,** (59), e2607.
49. Ruopp, M. D.; Perkins, N. J.; Whitcomb, B. W.; Schisterman, E. F., Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom J* **2008,** *50* (3), 419-30.
50. Freund, R. K.; Martin, B. J.; Jungschaffer, D. A.; Ullman, E. A.; Collins, A. C., Genetic differences in plasma corticosterone levels in response to nicotine injection. *Pharmacol Biochem Behav* **1988,** *30* (4), 1059-64.
51. Armario, A.; Montero, J. L.; Balasch, J., Sensitivity of corticosterone and some metabolic variables to graded levels of low intensity stresses in adult male rats. *Physiol Behav* **1986,** *37* (4), 559-61.
52. Kant, G. J.; Leu, J. R.; Anderson, S. M.; Mougey, E. H., Effects of chronic stress on plasma corticosterone, ACTH and prolactin. *Physiol Behav* **1987,** *40* (6), 775-9.
53. Jacobsen, K. R.; Kalliokoski, O.; Teilmann, A. C.; Hau, J.; Abelson, K. S., The effect of isoflurane anaesthesia and vasectomy on circulating corticosterone and ACTH in BALB/c mice. *Gen Comp Endocrinol* **2012,** *179* (3), 406-13.
54. Sundbom, R.; Jacobsen, K. R.; Kalliokoski, O.; Hau, J.; Abelson, K. S., Post-operative corticosterone levels in plasma and feces of mice subjected to permanent catheterization and automated blood sampling. *In Vivo* **2011,** *25* (3), 335-42.
55. Salas, M. A.; Evans, S. W.; Levell, M. J.; Whicher, J. T., Interleukin-6 and ACTH act synergistically to stimulate the release of corticosterone from adrenal gland cells. *Clin Exp Immunol* **1990,** *79* (3), 470-3.
56. Berger, K.; Sander, M.; Spies, C. D.; Weymann, L.; Bühner, S.; Lochs, H.; Wernecke, K. D.; von Heymann, C., Profound haemodilution during normothermic cardiopulmonary bypass influences neither gastrointestinal permeability nor cytokine release in coronary artery bypass graft surgery. *Br J Anaesth* **2009,** *103* (4), 511-7.
57. Swartbol, P.; Pärsson, H.; Truedsson, L.; Sjöholm, A.; Norgren, L., Aortobifemoral surgery induces complement activation and release of interleukin-6 but not tumour necrosis factor-alpha. *Cardiovasc Surg* **1996,** *4* (4), 483-91.
58. Desborough, J. P., The stress response to trauma and surgery. *Br J Anaesth* **2000,** *85* (1), 109-17.

59. Leon, L. R.; Blaha, M. D.; DuBose, D. A., Time course of cytokine, corticosterone, and tissue injury responses in mice during heat strain recovery. *J Appl Physiol (1985)* **2006,** *100* (4), 1400-9.
60. Dalm, S.; Enthoven, L.; Meijer, O. C.; van der Mark, M. H.; Karssen, A. M.; de Kloet, E. R.; Oitzl, M. S., Age-related changes in hypothalamic-pituitary-adrenal axis activity of male C57BL/6J mice. *Neuroendocrinology* **2005,** *81* (6), 372-80.
61. Teilmann, A. C.; Kalliokoski, O.; Sørensen, D. B.; Hau, J.; Abelson, K. S., Manual versus automated blood sampling: impact of repeated blood sampling on stress parameters and behavior in male NMRI mice. *Lab Anim* **2014,** *48* (4), 278-91.
62. Nichols, D. J.; Chevins, P. F., Plasma corticosterone fluctuations during the oestrous cycle of the house mouse. *Experientia* **1981,** *37* (3), 319-20.
63. Atkinson, H. C.; Waddell, B. J., Circadian variation in basal plasma corticosterone and adrenocorticotropin in the rat: sexual dimorphism and changes across the estrous cycle. *Endocrinology* **1997,** *138* (9), 3842-8.
64. Bronson, F. H.; Desjardins, C., Endocrine responses to sexual arousal in male mice. *Endocrinology* **1982,** *111* (4), 1286-91.
65. Stanulis, E. D.; Matulka, R. A.; Jordan, S. D.; Rosecrans, J. A.; Holsapple, M. P., Role of corticosterone in the enhancement of the antibody response after acute cocaine administration. *J Pharmacol Exp Ther* **1997,** *280* (1), 284-91.
66. Dominguez, G.; Faucher, P.; Henkous, N.; Krazem, A.; Piérard, C.; Béracochéa, D., Stress induced a shift from dorsal hippocampus to prefrontal cortex dependent memory retrieval: role of regional corticosterone. *Front Behav Neurosci* **2014,** *8*, 166.
67. Bernatova, I.; Key, M. P.; Lucot, J. B.; Morris, M., Circadian differences in stress-induced pressor reactivity in mice. *Hypertension* **2002,** *40* (5), 768-73.
68. Chen, L.; Lu, D.; Sun, K.; Xu, Y.; Hu, P.; Li, X.; Xu, F., Identification of biomarkers associated with diagnosis and prognosis of colorectal cancer patients based on integrated bioinformatics analysis. *Gene* **2019,** *692*, 119-125.
69. Cai, Y. L.; Lin, Y. X.; Jiang, L. S.; Ye, H.; Li, F. Y.; Cheng, N. S., A Novel Nomogram Predicting Distant Metastasis in T1 and T2 Gallbladder Cancer: A SEER-based Study. *Int J Med Sci* **2020,** *17* (12), 1704-1712.
70. Annunziata, M. A.; Muzzatti, B.; Bidoli, E.; Flaiban, C.; Bomben, F.; Piccinin, M.; Gipponi, K. M.; Mariutti, G.; Busato, S.; Mella, S., Hospital Anxiety and Depression Scale (HADS) accuracy in cancer patients. *Support Care Cancer* **2020,** *28* (8), 3921-3926.
71. Kim, J.; Bamlet, W. R.; Oberg, A. L.; Chaffee, K. G.; Donahue, G.; Cao, X. J.; Chari, S.; Garcia, B. A.; Petersen, G. M.; Zaret, K. S., Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. *Sci Transl Med* **2017,** *9* (398).
72. Jirkof, P.; Cesarovic, N.; Rettich, A.; Nicholls, F.; Seifert, B.; Arras, M., Burrowing behavior as an indicator of post-laparotomy pain in mice. *Front Behav Neurosci* **2010,** *4*, 165.
73. Leung, V. S. Y.; Benoit-Biancamano, M. O.; Pang, D. S. J., Performance of behavioral assays: the Rat Grimace Scale, burrowing activity and a composite behavior score to identify visceral pain in an acute and chronic colitis model. *Pain Rep* **2019,** *4* (2), e718.
74. Ray, M. A.; Johnston, N. A.; Verhulst, S.; Trammell, R. A.; Toth, L. A., Identification of markers for imminent death in mice used in longevity and aging research. *J Am Assoc Lab Anim Sci* **2010,** *49* (3), 282-8.
75. Tsuji, Y.; Koga, S.; Ibayashi, H.; Nose, Y.; Akazawa, K., Prediction of the prognosis of liver cirrhosis in Japanese using Cox's proportional hazard model. *Gastroenterol Jpn* **1987,** *22* (5), 599-606.
76. Puleo, F.; Nicolle, R.; Blum, Y.; Cros, J.; Marisa, L.; Demetter, P.; Quertinmont, E.; Svrcek, M.; Elarouci, N.; Iovanna, J.; Franchimont, D.; Verset, L.; Galdon, M. G.; Devière, J.; de Reyniès, A.; Laurent-Puig, P.; Van Laethem, J. L.; Bachet, J. B.; Maréchal, R., Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. *Gastroenterology* **2018,** *155* (6), 1999-2013.e3.

77. Shi, H.; Wei, Y.; Cheng, S.; Lu, Z.; Zhang, K.; Jiang, K.; Xu, Q., Survival prediction after upfront surgery in patients with pancreatic ductal adenocarcinoma: Radiomic, clinic-pathologic and body composition analysis. *Pancreatology* **2021,** *21* (4), 731-737.
78. 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* **2017,** *12* (6), e0179588.
79. Herrmann, K.; Flecknell, P., Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals: A retrospective review. *Altern Lab Anim* **2018,** *46* (5), 273-289.
80. Mei, J.; Banneke, S.; Lips, J.; Kuffner, M. T. C.; Hoffmann, C. J.; Dirnagl, U.; Endres, M.; Harms, C.; Emmrich, J. V., Refining humane endpoints in mouse models of disease by systematic review and machine learning-based endpoint definition. *Altex* **2019,** *36* (4), 555-571.
81. Canbay, A.; Feldstein, A. E.; Higuchi, H.; Werneburg, N.; Grambihler, A.; Bronk, S. F.; Gores, G. J., Kupffer cell engulfment of apoptotic bodies stimulates death ligand and cytokine expression. *Hepatology* **2003,** *38* (5), 1188-98.
82. Muriel, P.; Escobar, Y., Kupffer cells are responsible for liver cirrhosis induced by carbon tetrachloride. *J Appl Toxicol* **2003,** *23* (2), 103-8.
83. Ramadori, G.; Knittel, T.; Odenthal, M.; Schwögler, S.; Neubauer, K.; Meyer zum Büschenfelde, K. H., Synthesis of cellular fibronectin by rat liver fat-storing (Ito) cells: regulation by cytokines. *Gastroenterology* **1992,** *103* (4), 1313-21.
84. Arias, M.; Sauer-Lehnen, S.; Treptau, J.; Janoschek, N.; Theuerkauf, I.; Buettner, R.; Gressner, A. M.; Weiskirchen, R., Adenoviral expression of a transforming growth factor-beta1 antisense mRNA is effective in preventing liver fibrosis in bile-duct ligated rats. *BMC Gastroenterol* **2003,** *3*, 29.
85. Kim, K. H.; Kim, H. C.; Hwang, M. Y.; Oh, H. K.; Lee, T. S.; Chang, Y. C.; Song, H. J.; Won, N. H.; Park, K. K., The antifibrotic effect of TGF-beta1 siRNAs in murine model of liver cirrhosis. *Biochem Biophys Res Commun* **2006,** *343* (4), 1072-8.
86. Reinehr, R.; Becker, S.; Keitel, V.; Eberle, A.; Grether-Beck, S.; Häussinger, D., Bile salt-induced apoptosis involves NADPH oxidase isoform activation. *Gastroenterology* **2005,** *129* (6), 2009-31.
87. Unsal, V.; Cicek, M.; Sabancilar, İ., Toxicity of carbon tetrachloride, free radicals and role of antioxidants. *Rev Environ Health* **2021,** *36* (2), 279-295.

8. Acknowledgement

I write the acknowledgment with great gratitude. On the occasion of the completion of my dissertation, I would like to express my sincere gratitude to everyone, who has supported, cared and helped me.

I would like to express my sincere thanks, dear Prof. Dr. med. Brigitte Vollmar, for giving me the opportunity to work and study at our institute, the Rudolf-Zenker-Institut für Experimentelle Chirurgie. Thank you for your suggestions for my publications, and also thank you for your support completing this dissertation.

I would like to express my deepest thanks to my supervisor, dear PD. Dr. Dietmar Zechner. Thank you for your patient and effective guidance in my studies. Without your support, I could not have been able to publish my papers and complete this dissertation. You are a kind gentleman, it is the lucky for me to be your student.

I would like to express my sincere gratitude to my good friend, brother, classmate, colleague and teacher, Dr. Xianbin Zhang. Thank you for your support and help in my study, work and life. Words cannot express my gratitude. May our friendship last forever.

I would like to thank Dr. Simone Kumstel, Nico Seume, Wiebke Nierath, Dr. Wentao Xie and Ahmed Abdelrahman. Thank you very much for your friendship and support.

Many thanks to Dr. Kerstin Abshagen for guiding me, when performing the bile duct ligation. Many thanks also to my dear friend Anja Gellert, for your friendship and administrative support.

I would like to thank Frau Berit Blendow, Frau Dorothea Frenz, Frau Eva Lorbeer, and Frau Maren Nerowski for your excellent technical assistance.

I would like to thank my dear country, China. Thank you very much for covering the cost of my living expenses in Germany. This supported me in Germany and helped me to focus on my studies.

Finally, I would like to thank my family, especially my wife，Yue Cui. Your love and tolerance supported me when working on my dissertation.

9. List of abbreviations

Hypothalamus-pituitary-adrenal	HPA
Corticotrophin-releasing hormone	CRH
Adrenocorticotrophic hormone	ACTH
Bile duct ligation	BDL
Carbon tetrachloride	CCl_4
Aspartate aminotransferase	AST
Alanine aminotransferase	ALT
Receiver operating characteristic curve	ROC curve
Harrell's concordance index	C-indices
Support vector machine	SVM
Hazard ratio	HR
Area under the curve	AUC
Transforming growth factor b1	TGF-b1
Nicotinamide adenine dinucleotide phosphate	NADPH

10. Eidesstattliche Versicherung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und nur unter der Verwendung der angegebenen Quellen und Hilfsmittel erstellt habe. Die aus anderer Literatur verwendeten Inhalte sind als solche kenntlich gemacht. Die Regeln zur Sicherung guter wissenschaftlicher Praxis wurden beachtet. Ich versichere, dass ich für die inhaltliche Erstellung der vorliegenden Arbeit nicht die entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen habe. Ich bestätige, dass diese Arbeit an keiner anderen als der Universität Rostock zur Erlangung des Grades Doktor der Medizin vorgelegt wurde.

Rostock, 24.09.2023 ……..

Ort, Abgabedatum Vollständige Unterschrift

11. Thesen

- Overwhelming distress of laboratory animals is not only inconsistent with the requirements of animal ethics, but may also lead to unreliable scientific conclusions and affect reliability and reproducibility of animal experiments.
- Liver fibrosis is a pathological feature in a variety of liver diseases and often proceeds to liver cirrhosis and failure. Although mouse models of liver fibrosis, such as bile duct ligation and intoxication by CCl_4, are widely used in medical research, there are few available reports on distress evaluation in these animal models.
- In order to assess the welfare of animals, it is important to evaluate, which methods can measure distress with high sensitivity and specificity for specific animal models.
- Often a distress score, spontaneous behavior (burrowing and nesting activity), physiology (the body weight), and plasma corticosterone is assessed, when judging distress of animals.
- The present study demonstrated that the peak and duration of corticosterone kinetics can be used to evaluated distress in mouse models of liver damage.
- Several non-invasive parameters, such as a distress score, the body weight and burrowing activity, also proved to be useful when judging distress (evaluated by ROC curves) in mouse models of liver fibrosis.
- Multivariate analysis using Support Vector classification combined these parameters, and, thereby, allowed us to compare distress of two distinct models of liver fibrosis.
- This dissertation demonstrated that bile duct ligation causes much more distress than CCl_4-induced liver fibrosis.
- This dissertation also suggests that a weight loss of 10-20% combined with a reduction in burrowing activity of more than 79.4% can predicted the death of animals within two days in a murine model for cholestasis.
- Thus, the present study provided new ideas and approaches for evaluating and comparing the distress of animal models, and determined early humane endpoints for timely euthanasia.
- These results provide a solid basis for reducing distress in future animal experiments and point the way for improving the quality of preclinical research.

12. Curriculum vitae

PERSONAL DETAILS

Name:	Guanglin Tang
Email:	guanglintang@hotmail.com
Address:	Linjiang East Road 53, Fushun, China
Date/place of birth:	09.01.1988 in Liaoning, China

CURRENT POSITION

Since 10/2020	Surgical doctor, Department of General Surgery, Fushun Central Hospital, China

EDUCATION

12/2017-10/2020	MD student at the Institute of Experimental Surgery, University Rostock, Germany
10/2007- 07/2014	Bachelor and Master of Medicine in Surgery, Dalian Medical University, China

WORK EXPERIENCE

08/2014 - 11/2017	Department of General Surgery, Fushun Central Hospital, China

GRANT

Chinese Government Graduate Student Overseas Study Program. Ministry of Education of the People's Republic of China (NO. 201808080167).

PUBLICATIONS

1. Kumstel S, Tang G, Zhang X, Kerndl H, Vollmar B, Zechner D. Grading Distress of Different Animal Models for Gastrointestinal Diseases Based on Plasma Corticosterone Kinetics. Animals (Basel). 2019 Apr 3;9(4):145.
2. Tang G, Seume N, Häger C, Kumstel S, Abshagen K, Bleich A, Vollmar B, Talbot SR, Zhang X, Zechner D. Comparing distress of mouse models for liver damage. Sci Rep. 2020 Nov 13;10(1):19814.
3. Zhang X[#], Kumstel S[#], Tang G[#], Talbot SR, Seume N, Abshagen K, Vollmar B, Zechner D. A rational approach of early humane endpoint determination in a murine model for cholestasis. ALTEX. 2020;37(2):197-207. [#]equal contribution
4. Talbot SR, Kumstel S, Schulz B, Tang G, Abdelrahman A, Seume N, Wendt EHU, Eichberg J, Häger C, Bleich A, Vollmar B, Zechner D. Robustness of a multivariate composite score when evaluating distress of animal models for gastrointestinal diseases.

Sci Rep. 2023 Feb 14;13(1):2605.

5. Zechner D, Schulz B, Tang G, Abdelrahman A, Kumstel S, Seume N, Palme R, Vollmar B. Generalizability, Robustness and Replicability When Evaluating Wellbeing of Laboratory Mice with Various Methods. Animals (Basel). 2022 Oct 25;12(21):2927.
6. Tang G, Nierath WF, Palme R, Vollmar B, Zechner D. Analysis of Animal Well-Being When Supplementing Drinking Water with Tramadol or Metamizole during Chronic Pancreatitis. Animals (Basel). 2020 Dec 5;10(12):2306.
7. Kumstel S, Janssen-Peters H, Abdelrahman A, Tang G, Xiao K, Ernst N, Wendt EHU, Palme R, Seume N, Vollmar B, Thum T, Zechner D. MicroRNAs as systemic biomarkers to assess distress in animal models for gastrointestinal diseases. Sci Rep. 2020 Oct 9;10(1):16931.

13. Appendix

The following publications are part of the Appendix:

Study I:
Kumstel S, **Tang G**, Zhang X, Kerndl H, Vollmar B, Zechner D. Grading Distress of Different Animal Models for Gastrointestinal Diseases Based on Plasma Corticosterone Kinetics. Animals (Basel). 2019 Apr 3;9(4):145. (IF 2022: 3.0)

Study II:
Tang G, Seume N, Häger C, Kumstel S, Abshagen K, Bleich A, Vollmar B, Talbot SR, Zhang X, Zechner D. Comparing distress of mouse models for liver damage. Sci Rep. 2020 Nov 13;10(1):19814. (IF 2022: 4.6)

Study III:
Zhang X[#], Kumstel S[#], **Tang G[#]**, Talbot SR, Seume N, Abshagen K, Vollmar B, Zechner D. A rational approach of early humane endpoint determination in a murine model for cholestasis. ALTEX. 2020;37(2):197-207. (IF 2022: 5.6)
[#]equal contribution

animals

MDPI

Article

Grading Distress of Different Animal Models for Gastrointestinal Diseases Based on Plasma Corticosterone Kinetics

Simone Kumstel *, Guanglin Tang, Xianbin Zhang, Hagen Kerndl, Brigitte Vollmar and Dietmar Zechner

Rudolf-Zenker-Institute of Experimental Surgery, University Medical Center, 18057 Rostock, Germany; guanglintang@hotmail.com (G.T.); zhangxianbin@hotmail.com (X.Z.); hagen.kerndl@uni-rostock.de (H.K.); brigitte.vollmar@uni-rostock.de (B.V.); dietmar.zechner@uni-rostock.de (D.Z.)

* Correspondence: simone.kumstel@uni-rostock.de; Tel.: +49-381-494-2512

Received: 13 February 2019; Accepted: 31 March 2019; Published: 3 April 2019

Simple Summary: Animal welfare is an important aspect of biomedical research. Many regulations have been implemented to combine high quality of research with minimal harm to laboratory animals. These guidelines also demand a prospective severity assessment of each animal model. A comparison of distress between animal models could allow realistic harm and benefit analysis and an appropriate use of refinement methods. However, studies comparing distress between different animal models are still rare. One good parameter for analyzing distress is the concentration of the stress hormone corticosterone in the blood. Therefore, we compared the corticosterone kinetics of distinct gastrointestinal animal models. The aim of this study was to evaluate which parameter the highest corticosterone concentration or the duration of increased stress hormone level could be used to quantify distress. We observed a significant increase of corticosterone 30 min after stress induction in all animal models. However, the corticosterone kinetics differed between the distinct interventions. Both the absolute value and the duration of increased corticosterone level correlated directly with an assessed distress score. We conclude that both variables of corticosterone kinetics are valid parameters to compare distress between animal models.

Abstract: Comparative studies for evaluating distress in established animal models are still rare. However, this issue is becoming more important as a consequence of worldwide appreciation of animal welfare. One good parameter for evaluating distress is the quantification of corticosterone. We hypothesized that not just the absolute value but also the duration of increased corticosterone concentration in the blood is an important aspect for evaluating animal distress. Therefore, we analyzed plasma corticosterone concentrations 30, 60, 120, and 240 min after induction of pancreatitis by cerulein, liver damage by carbon tetrachloride, liver damage by bile duct ligation, and after orthotopic injection of pancreatic cancer cells. We also evaluated corticosterone kinetics after injection of distinct carrier substances. Compared to phosphate buffered saline, dimethyl sulfoxide leads to dose-dependent higher and longer-lasting circulating corticosterone concentrations. In all disease models, we observed significantly increased corticosterone concentration 30 min after stress induction. However, the corticosterone kinetics differed among the animal models. Both the absolute value of corticosterone concentration and the duration correlated positively with the quantification of animal distress by a score sheet. This suggests that both variables of corticosterone kinetics might provide a solid basis for comparing and grading distress of different animal models.

Keywords: corticosterone; stress; pancreatitis; liver fibrosis; pancreatic cancer; hypothalamic-pituitary-adrenal axis

1. Introduction

Animal welfare is important for biomedical research and many regulations have been implemented over the last decades to combine high quality of research with minimal harm to animals. In many countries and especially in the European Union (EU) after the introduction of the EU directive 63/2010, animal-based research should strictly follow the 3R principle (reduce, refine, replace) and the severity level of each animal experiment must be defined and reported by the scientists. In the last decades, many animal models for different diseases have been developed for biomedical research. However, usually only parameters describing the pathology are evaluated and studies comparing the distress between animal models are still rare [1,2].

For the evaluation of distress in laboratory animals, the analysis of plasma corticosterone has proven to be a sensitive method [3–5]. However, its validity for comparing distress between different animal models and strains still needs to be explored. Corticosterone is the most important glucocorticoid regulating the stress response in rodents. Different stressors induce the activation of the hypothalamic–pituitary-adrenal (HPA) axis and provoke the release of corticotrophin-releasing hormone (CRH) and arginine vasopressin (AVP) from neurons in the paraventricular nucleus of the hypothalamus (PVN). The transport of these factors to the pituitary gland leads to the secretion of adrenocorticotrophic hormone (ACTH), which stimulates the adrenal cortex to release corticosterone into the circulation [6]. Corticosterone regulates the metabolism to cover the energy supply for the stress response [7] and modulates defense reactions to stress [4].

The plasma corticosterone level, however, does not always indicate distress, but is also influenced by physiological processes such as the circadian rhythm [3,8,9], sexual arousal [10], and the estrus cycle [3,11,12]. Nevertheless, it is widely accepted that repeated and unpredictable stressors increase the corticosterone concentration in the blood [3,5,13]. However, the kinetics of corticosterone concentration in the blood were reported to be distinct between different stressors. For example, timing differences in the up- and downregulation of corticosterone in the blood were observed after restraint stress, electric shocks, shaker stress, heat stress, and forced swimming [3,14–16]. These studies suggest that an assessment of corticosterone concentration at only one point in time is insufficient, for a conclusively comparison of distress between different stressors. This point is especially valid when comparing diseases in distinct animal models, which are induced by completely different methods, for instance by injection of chemicals or by surgical intervention. The purpose of this study was to analyze the corticosterone kinetics after inducing distinct gastrointestinal diseases. Gastrointestinal diseases are the most common cause of hospital admission and the third most common cause of death [17]. In particular, pancreatic cancer, pancreatitis, and liver fibrosis require new treatment strategies [18–20]. Therefore, these diseases are part of many preclinical trials using animal models [21–23]. In this explorative study, we assessed the optimal time point to compare the maximal plasma corticosterone concentration after induction of different gastrointestinal animal models using different mouse strains. Additionally, we evaluated which parameter of the corticosterone profile the absolute value or the duration of increased corticosterone concentration is a good indicator of distress. A comparison of distress of these widely-used gastrointestinal animal models could be used to adjust refinement methods in future.

2. Materials and Methods

2.1. Study Background, Animals, and Interventions

The current study is part of a DFG (German Research Foundation)-funded multicenter approach (FOR 2591) including 15 different research groups of eight institutions in Germany and Switzerland, involving five different animal species. The overall goal of this research group is to establish a species and animal model severity assessment framework and consequently to minimize severity in laboratory animals by employing improved refinement methods [24]. The current study is therefore an essential first step to analyze and to compare corticosterone kinetics after different interventions and to evaluate the optimal time point to measure corticosterone.

The mice were kept in a room with 12/12 h dark light cycle (light period: 07.00 h–19.00 h) with food and water ad libitum. The mice were kept in groups of 2–5 animals in type III cages and enrichment was provided by paper role and rodent wood. The mice were bred in our own facility under specific pathogen free conditions. All experiments were approved by the Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei Mecklenburg-Vorpommern (7221.3-1-019/15; 7221.3-1-002/17). The decisions of the local authority are in accordance with the Protection of Animals Act (Deutsches Tierschutz Gesetz) and the European Directive 2010/63/EU. For all animal models, male mice were used, to exclude the influence of the estrus cycle on the corticosterone concentration [3,12,13]. During all procedures, the mice were picked up by their tail. For analyzing the distress after injecting distinct solvents, which we usually use as different therapeutic agents, 15–20-week-old C57BL/6J mice were restrained for about 10–15 seconds while injecting once 2.5 µL/g phosphate buffered saline (PBS), and a low volume (1.25 µL/g) or high volume (2.5 µL/g) of 100% dimethyl sulfoxide (DMSO, Merck, Darmstadt, Germany) into the peritoneal cavity.

2.2. Chemical Induction of Pancreatitis and Liver Damage

Pancreatitis was induced in male 17–19 weeks old C57BL/6J mice by three intraperitoneal (i.p.) injections of cerulein (50 µg/kg; Merck, Darmstadt, Germany) diluted in 0.9% sodium chloride, at a rate of one every hour, three times a week up to 35 days (thus Monday, Wednesday and Friday). The blood for the corticosterone analysis was collected after the third cerulein injection on the second day of cerulein administration. This procedure mimics the early phase of chronic pancreatitis [25]. In order to induce liver damage by carbon tetrachloride (CCl4), male mice of the BALB/cANCrl strain were used at the age of 17–20 weeks. This liver fibrosis model is well established in this mouse strain at our institute [26]. Twice a week (Monday and Friday) for up to 42 days, CCl4 (Merck, Darmstadt, Germany) was i.p. injected at a dose of 0.25 µL/g body weight (after 4-fold dilution in corn oil). This procedure mimics the early phase of an animal model for liver fibrosis [26]. The blood for the analysis of corticosterone was taken on the second day of CCl4 injection. Starting on the day before the first injection, analgesia was provided continuously for both animal models by adding 1250 mg/L metamizol (Ratiopharm, Ulm, Germany) to the drinking water until the end of the experiment.

2.3. Surgical Induction of Liver Damage and Pancreatic Cancer

The bile duct ligation (BDL) was performed on male BALB/cANCrl mice at the age of 16–20 weeks. The mice were anesthetized with 1.2–2.5% isoflurane and 5 mg/kg carprofen (Rimadyl®, Pfizer GmbH Berlin, Germany) was injected subcutaneously (s.c.) for perioperative analgesia. The eyes were kept wet by eye ointment. A midline laparotomy was executed and the common bile duct was ligated three times with 5-0 silk and transected between the two most distal ligations. The peritoneum and the skin were closed separately by 5-0 prolene suture (Johnson & Johnson MEDICAL GmbH, New Brunswick, NJ, USA) and the mice were placed in front of a heating lamp. For this procedure, the mice were anaesthetized for 25–35 minutes.

For the orthotopic injection of cancer cells, male 17–24-week-old C57BL/6J mice were used, since this model is well established in this mouse strain [21]. The mice were anesthetized with 1.2–2.5% isoflurane and 5 mg/kg carprofen (Rimadyl®, Pfizer GmbH) were applied by s.c. injection for analgesia. The eyes were kept wet by eye ointment. The abdomen of the mice was shaved and disinfected and the abdominal cavity was opened by laparotomy and 5 µL of the cell suspension (murine cell line 6606 PDA, 2.5 × 105/5 µL cells in matrigel) were injected slowly with a 25-µL syringe into the head of the pancreas (Hamilton Syringe, Reno, NV, USA). The pancreas was placed back into the cavity and the peritoneum was closed by a coated 5-0 vicryl suture (Johnson & Johnson MEDICAL GmbH). The skin was sewed with a 5-0 prolene suture (Johnson & Johnson MEDICAL GmbH). The surgery lasted 15–25 min for each of the mice.

2.4. Preparation of Blood Samples, Evaluation of Distress Score, and Analysis of Plasma

The mice were assigned to four groups in a haphazard manner (no randomization was performed). The blood sampling was executed at the indicated time points (30, 60, 120, and 240 min) after the different i.p. injections or the awakening of the mice after the surgical interventions. The blood was collected only once from each mouse. To assess the plasma corticosterone levels without stress induction (pre-values), the blood of each mouse was collected 2–4 weeks before distinct interventions were conducted. The mice were anesthetized with 5% isoflurane and 100–150 µL blood were collected within 3 minutes by retro orbital puncture in EDTA-Tubes (Microvette®, SARSTEDT, Nürnbrecht, Germany) to avoid an increase of corticosterone caused by the procedure [27,28]. The blood sampling was always carried out during a time period from 11:00 to 15:30 h, to avoid the circadian increase of the plasma corticosterone, which starts after 16:00 h (2 h before the dark phase) [3,8,9]. After the blood sampling and the surgical procedures, the mice were euthanized directly via cervical dislocation under isoflurane anesthesia to prevent a distress accumulation from both interventions. The tubes were centrifuged (1200× *g*, 10 min, 20 °C) to separate the plasma and the samples were stored at −20 °C. Immediately before blood collection, the distress score was assessed by only one person in a not blinded manner, according to a scoring sheet (Supplemental Tables S1 and S2), which was modified after Paster et al. [29].

The corticosterone concentration in the plasma samples was measured in a blinded fashion with the ELISA-Kit (DE 4164, Demeditec Diagnostics GmbH, Erfurt, Germany) following manufacturer instructions. To evaluate acinar cell damage, the activity of lipase in the blood plasma was analyzed in a blinded fashion with a photometer (cobas c111, Roche Diagnostics, Rotkreuz, Switzerland). As parameters for the acute liver damage the plasma activity of the aspartate aminotransferase (AST) and alanine aminotransferase (ALT) was measured with the identical photometer.

2.5. Data Analysis

All data were analyzed with the program SigmaPlot 12.0 (SYSTAT Software Inc., San Jose, USA). All graphs present data as box plots the 10th and 90th percentile as whiskers. The significances of differences were evaluated by Mann–Whitney rank-sum test, followed by Bonferroni correction. Non-parametric tests were used, because of the low power of normality tests at a limited sample size from 3–5 [30]. Differences with $p \leq 0.05$, divided by the number of comparisons to the pre-value ($p \leq 0.0125$), were considered to be significant. Correlations were described by calculating the Spearman correlation coefficient (number of samples: 17).

3. Results

3.1. Injection of Distinct Solvents Provokes Specific Corticosterone Profiles

Injection of PBS (2.5 µL/g) significantly increased the plasma corticosterone concentration at 30 min, followed by a rapid reduction of the stress hormone concentration (Figure 1A). However, after the i.p. injection of a low volume of DMSO (1.25 µL/g), the corticosterone concentration was significantly increased 30 min as well as 60 min after stress induction (Figure 1B). After the administration of a high DMSO volume (2.5 µL/g), we noticed a significantly increased corticosterone concentration 30 min, 60 min, and 120 min after stress induction (Figure 1C). The maximum of corticosterone concentration was detected at 120 min (Figure 1C). Even 4 hours after the injection, the corticosterone level remained slightly higher in comparison to the values without stress induction (Figure 1C). Independent of the corticosterone profile, the administration of all solvents led to a significant increase of this stress hormone 30 min after injection (Figure 1D).

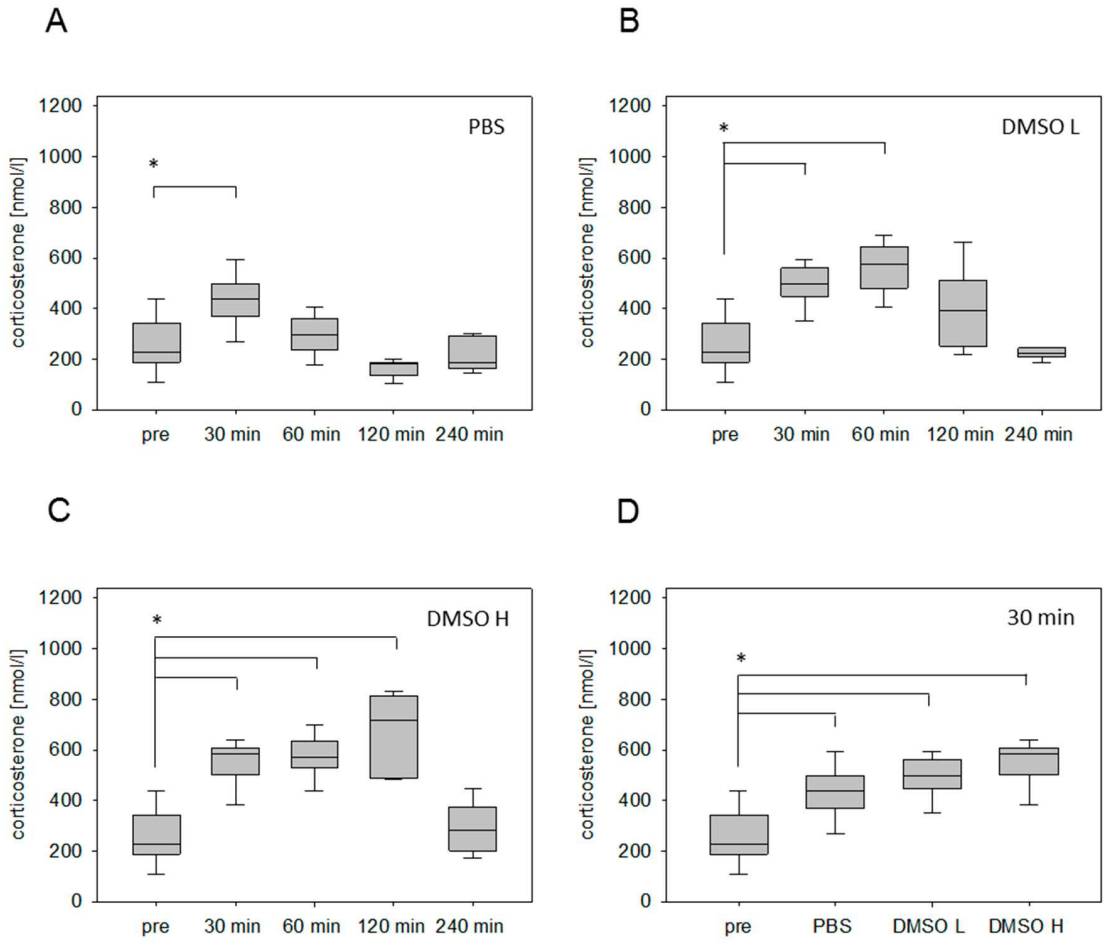


Figure 1. Corticosterone concentration in blood plasma after intraperitoneal (i.p.) injection of different solvents. Time-dependent glucocorticoid response before (pre) and after injection of phosphate buffered saline (PBS) (2.5 µL/g) (**A**), low volume of dimethyl sulfoxide (DMSO L: 1.25 µL/g) (**B**) and high volume of DMSO (DMSO H: 2.5 µL/g) (**C**). Comparison of corticosterone concentration measured 30 min after injection of indicated solvents (**D**). Significant differences: * $p \leq 0.0125$; $n = 20$ (pre); $n = 5$ (30, 60, 120, 240 min).

3.2. Intervention-Specific Profiles in Widely Used Gastrointestinal Mouse Models

The stress hormone response was also measured in an animal model for pancreatitis. The corticosterone concentration was significantly increased at 30 min and 60 min after cerulein injection (Figure 2A). The maximum was detected at 30 min followed by a constant reduction of the corticosterone concentration (Figure 2A). As a parameter for pancreatic tissue damage, lipase activity was quantified. The activity of this enzyme was significantly increased at 30 min, 60 min, 120 min, and 240 min (Figure 2B). We also analyzed the corticosterone profile after cancer cell injection into the pancreas in C57BL/6J mice. The corticosterone concentration was significantly increased at 30 min, 60 min, and 120 min after anesthesia (Figure 3A). The circulating stress hormone remained on a similar constant level until 120 min (Figure 3A). To measure tissue damage during cancer cell injection, lipase activity was evaluated at the distinct time points (Figure 3B). A slight increase of plasma lipase activity was detected at 60 min after cancer cell injection. However, all observed differences were not significant (Figure 3B). In BALB/c mice with CCl4-induced liver damage the highest concentration of corticosterone was observed at 30 min (Figure 4A). However, also after 60 min and 120 min the corticosterone concentration was significantly higher when compared to healthy animals (Figure 4A). In order to compare the stress hormone level with parameters for tissue damage, we determined the time course of plasma activity of AST and ALT (Figure 4B,C). Starting 30 min after the CCl4 injection,

we observed a significant increase in the activity of both transaminases (Figure 4B,C). A significant increase in the activity of both enzymes was also observed 60, 120, and 240 min after CCl4 injection (Figure 4B,C).

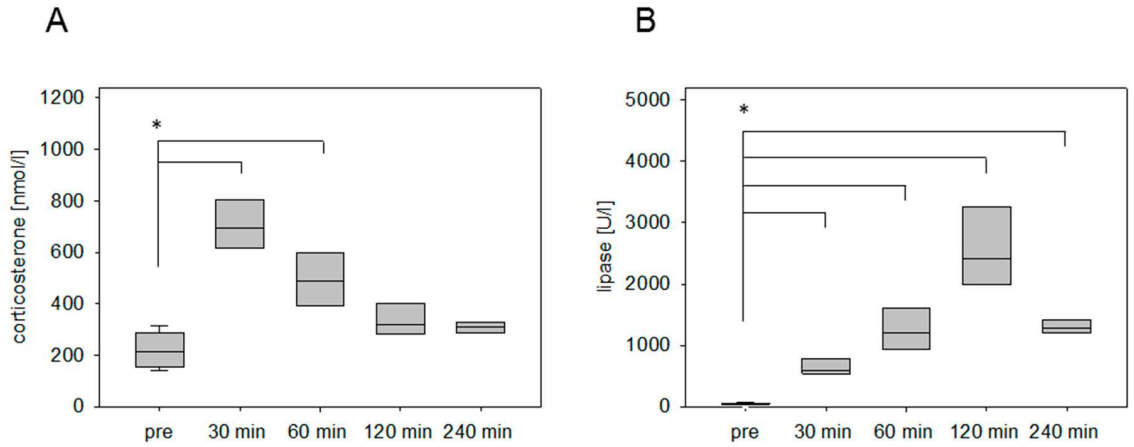


Figure 2. Corticosterone concentration and lipase activity after the induction of pancreatitis. Analysis of time-dependent corticosterone release before (pre) and after i.p. injection of cerulein (50 µg/kg) (**A**), as well as evaluation of lipase activity (**B**). Significant differences: * $p \leq 0.0125$; $n = 16$ (pre); $n = 4$ (30, 60, 120, 240 min).

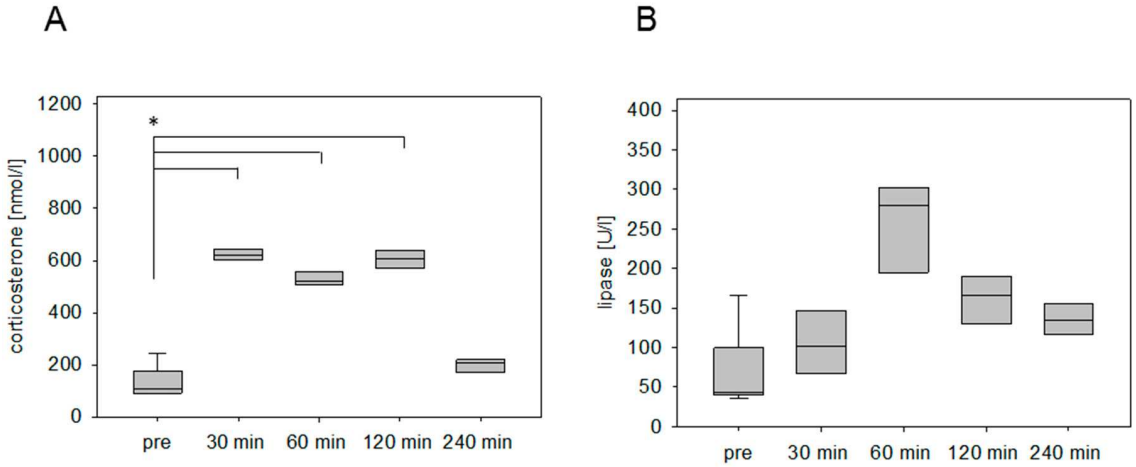


Figure 3. Corticosterone concentration after laparotomy and cancer cell injection into the pancreas. Evaluation of stress hormone release (**A**) as well as lipase activity (**B**) before surgery (pre) and at the indicated time points after cessation of anesthesia. Significant differences: * $p \leq 0.0125$; $n = 8$ (pre); $n = 4$ (30, 60, 120, 240 min).

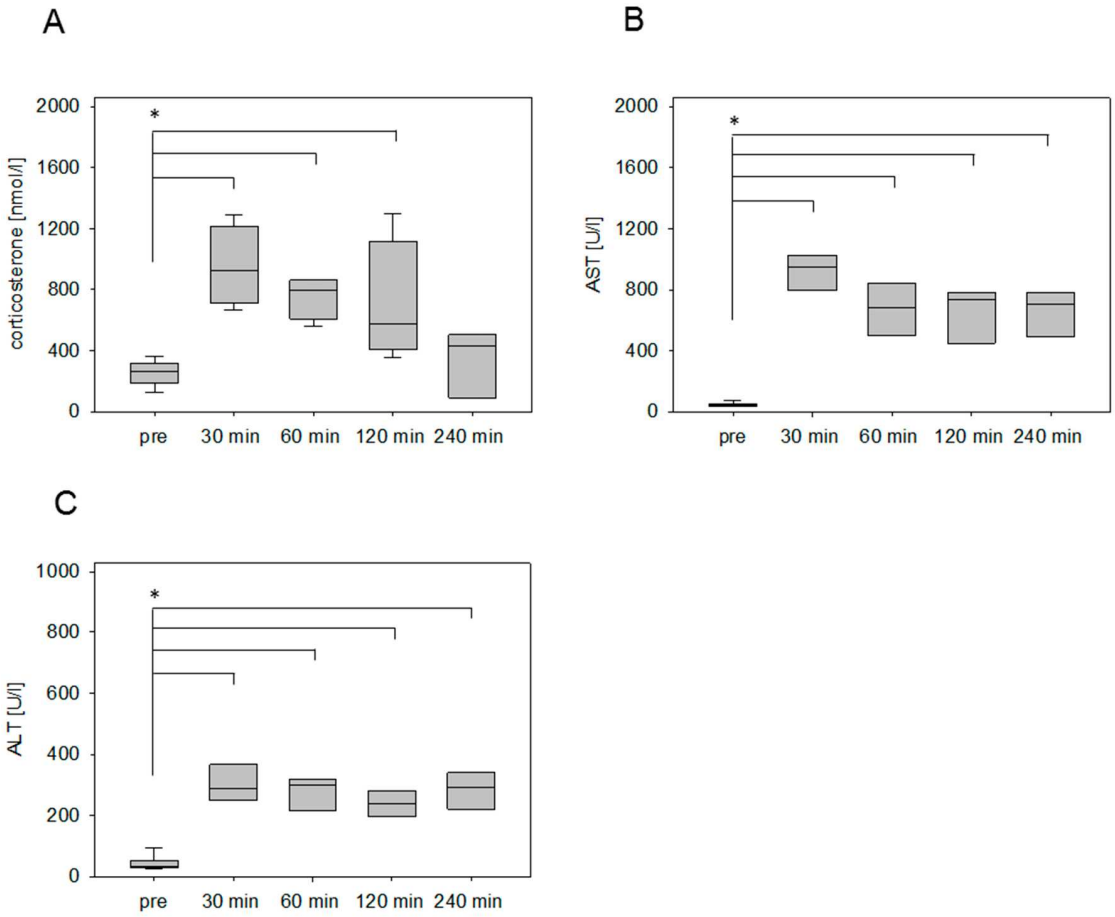


Figure 4. Corticosterone concentration and transaminase activities after chemical induction of liver damage. Analysis of stress hormone release of mice before (pre) and at the indicated time points after the i.p. injection of carbon tetrachloride (CCl4; 0.25 µL/g) (**A**) as well as the evaluation of aspartate aminotransferase (AST) (**B**) and alanine aminotransferase (ALT) activity (**C**). Significant differences: * $p \leq 0.0125$; $n = 15$ (pre), $n = 3$–4 (30, 60, 120, 240 min).

After the surgical induction of liver damage by bile duct ligation, a significantly increased concentration of circulating corticosterone was detected within 30 min after anesthesia (Figure 5A). In contrast to the chemically induced animal models (pancreatitis, liver damage by CCl4), the concentration of the stress hormone remains steady at a high level until 120 minutes. Even after 4 hours, the corticosterone concentration was still significantly higher in comparison to the plasma samples collected without any intervention (Figure 5A). The AST and ALT values, however, rose significantly at 30 min and sharply increased over time (Figure 5B,C).

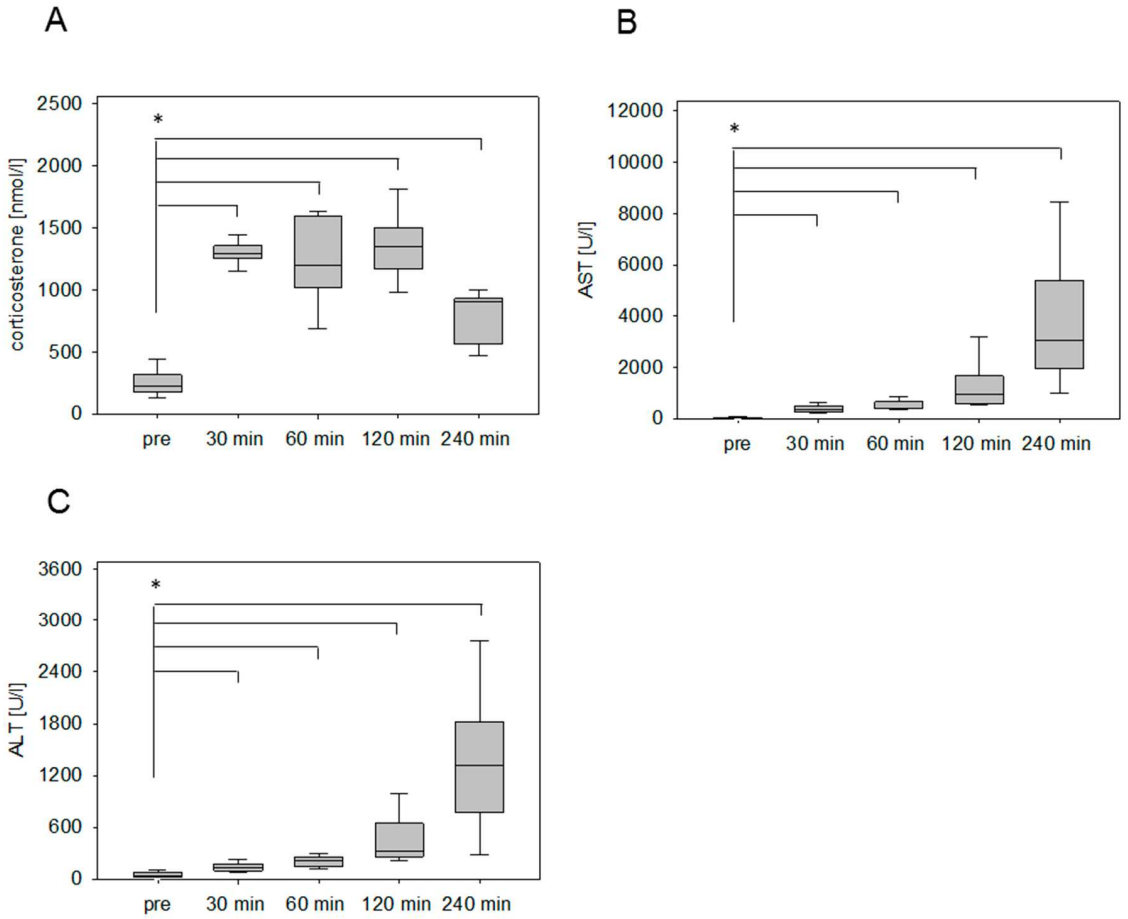


Figure 5. Stress hormone release and transaminase activities after surgical induction of liver damage by bile duct ligation (BDL). Evaluation of corticosterone concentration of mice before (pre) and after the awakening of the mice from anesthesia (**A**), as well as the evaluation of transaminase activity of AST (**B**) and ALT (C) as parameters for the tissue damage. Significant differences: $*p \leq 0.0125$; $n = 20$ (pre); $n = 5$ (30, 60, 120, 240 min).

3.3. Specific Distress Levels Caused by Distinct Gastrointestinal Diseases

In order to quantify distress, we evaluated the body weight, general health and the spontaneous as well as flight behavior of mice in all disease models as defined by a score sheet (for details on the score sheet see supplemental Tables S1 and S2). In the animal model for pancreatitis we observed 30 min after the last cerulein injection an abnormal posture (score 3) in one animal. All other mice received a score of 0. This resulted in an average distress score of 0.75 at the 30-min time point, whereas no distress was detected at the 60-min time point (Figure 6A). The injection of 6606PDA cells into the pancreas resulted in a significant increase of distress (score 2) 30 min after anesthesia, since all mice had a ruffled fur (Figure 6B). At 60 min no distress could be scored anymore (Figure 6B). In the animal model for liver fibrosis by CCl4 all mice had ruffled fur at the 30- and 60-min time point (Figure 6C). This resulted in an average distress score of 2 at both time points. The surgical induction of liver damage by BDL resulted in a very high distress score of 5, due to the accumulation of distress indicators including abnormal posture and passive behavior after touch (Figure 6D). At 60 min just abnormal posture (score 3) was still noticed in all mice (Figure 6D).

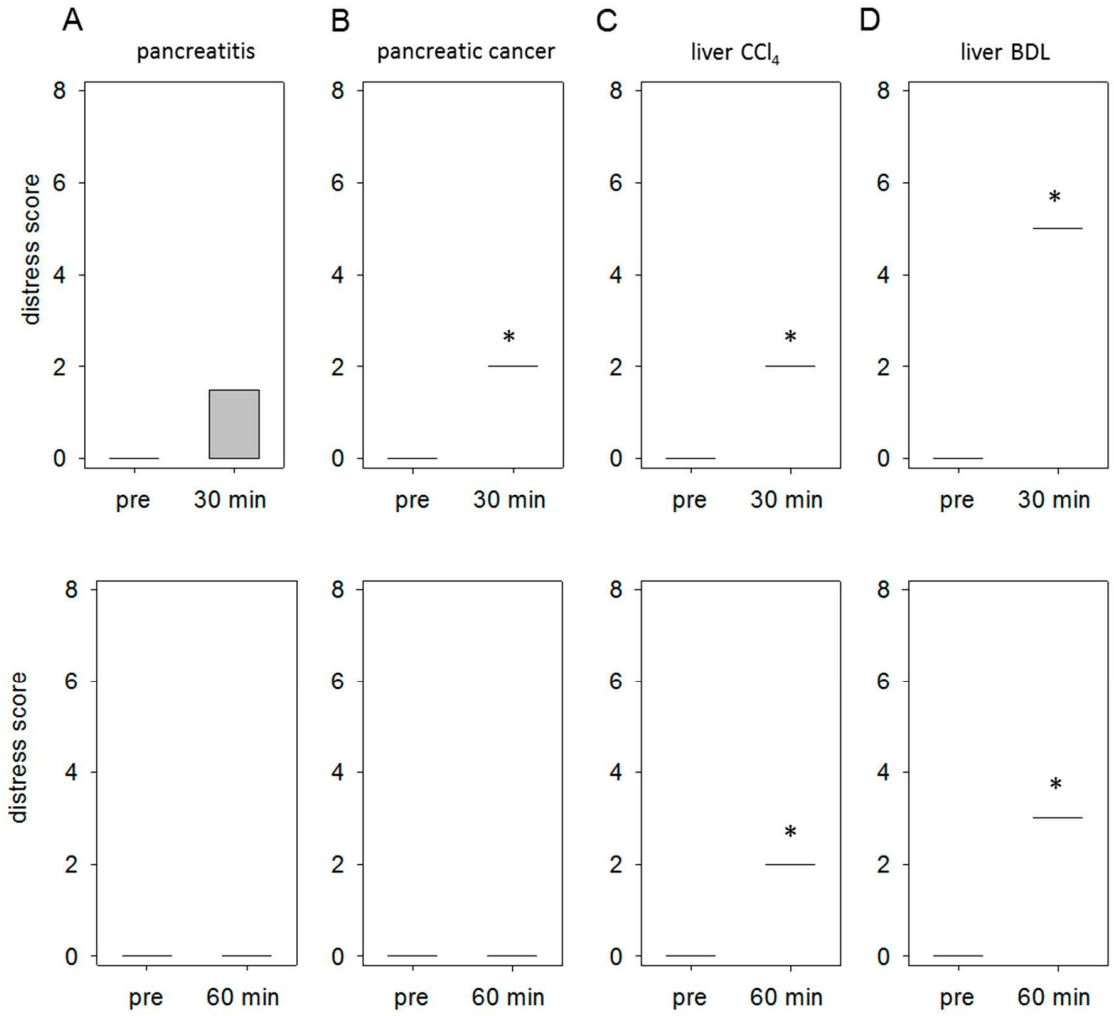


Figure 6. Comparison of distress between distinct mouse models. Comparison of the distress score at 30 min and 60 min after injection or end of anesthesia to the score taken before any intervention (pre). Thirty minutes after the induction of pancreatitis an abnormal posture was observed in only one animal, which results in an average distress score of 0.75 (**A**). Tumor cell injection into the pancreas resulted in a score of 2 (ruffled fur) in all mice at 30 min (**B**). Induction of liver damage by CCl4 led to a score of 2 (all animals displays ruffled fur) at 30 as well as at 60 min (**C**). Bile duct ligation resulted in a score of 5 at 30 min (due to abnormal posture and passive behavior after touch) and a score of 3 at 60 min (abnormal posture) (**D**). Significant differences: * $p \leq 0.05$; n = 4–5 (pre); n = 4–5 (30, 60 min).

4. Discussion

The present study demonstrates a significant increase of corticosterone concentration within 30 min independent of the stressor and the mouse strain (Figures 1–5). However, some interventions, such as injection of cancer cells into the pancreas, provoke a long lasting corticosterone response (Figure 5). Other interventions, such as the injection of cerulein, lead to a transient corticosterone response (Figure 2). Interestingly, not only the absolute value of corticosterone concentration but also the duration of the corticosterone response correlates well with the quantification of animal distress by a score sheet (Table 1).

Table 1. Animal models ranked according to distress, as well as absolute value and duration of corticosterone concentration. The mouse strains used for each animal model are specified. BDL: bile duct ligation.

BALB/c				**C57BL/6J**		
Distress (mean value of distress scored at 30 min)						
BDL	>	CCl_4	=	tumor cell injection	>	pancreatitis
5.0	>	2.0	=	2.0	>	0.75
Corticosterone concentration at 30 min (interquartile range 25–75% in nmol/L)						
BDL	>	CCl_4	>	tumor cell injection	~	pancreatitis
1218.1–1383.4	>	710.6–1215.1	>	599.0–649.6	~	602.4–831.6
Duration of corticosterone response (latest time point in minutes with a significant increase in corticosterone concentration)						
BDL	>	CCl_4	=	tumor cell injection	>	pancreatitis
240	>	120	=	120	>	60

The i.p. injection of different carrier substances is widely used in animal experiments for chemical disease induction and therapeutic intervention. Nevertheless, this intervention represents an important stressor, measurable by an increase of the stress hormone corticosterone. To evaluate the distress of the most important carrier substances of water- and lipo-soluble drugs, corticosterone kinetics after PBS and DMSO injection have been analyzed. A transient corticosterone response was noticed after PBS injection (Figure 1A). Other studies also demonstrated a quick downregulation of the stress hormone within 60 min [31,32]. This quick response is initiated by corticosterone itself due to a receptor-mediated negative feedback inhibition of the precursor hormones CRH and ACTH in the neurons of the PVN [33–35]. However, it is not possible to clearly differentiate between distress caused by handling of the animals and distress caused by the injection. The distress of only handling animals also leads to a significant increase in blood corticosterone concentration within 15 min, followed by a quick decrease, similar to observations made after PBS injection [36]. This suggests that PBS injection and handling of mice cause similar levels of distress. After DMSO administration, a longer-lasting corticosterone response in comparison to PBS injection was detected. Some studies suggest that DMSO can cause a longer-lasting activation of the HPA axis, independent of the nerval innervations of the hypothalamus [37,38]. We refrained from comparing the i.p. injections of PBS and DMSO directly with the induction of gastrointestinal disease, because the application of analgesia might have influenced the corticosterone level.

Intervention-specific corticosterone characteristics were analyzed after chemical and surgical disease inductions in liver and pancreas. Acute liver damage was induced in the murine liver of BALB/c mice by chemical induction with CCl4 and surgical intervention by BDL. Both interventions caused a significant increase of corticosterone concentration within 30 min. However, the injection of CCl4 caused a transient increase in corticosterone concentration, whereas the BDL provoked a significantly increased corticosterone concentration for a longer period (cf. Figures 4 and 5). Similar differences were observed when analyzing animal models, which focus on pancreatic diseases in the C57BL/6J strain. The injection of cerulein caused a transient increase in corticosterone, whereas the surgical induction of pancreatic cancer by laparotomy and cell injection in the pancreas caused a significant increase in corticosterone concentration for a longer period (cf. Figures 2 and 3). Even so, it is widely accepted that the absolute corticosterone concentration differs among distinct mice strains [5,15,32], the above-mentioned characteristic kinetics after chemical and surgical disease induction could be observed in both strains. A long-lasting corticosterone response for 2–6 hours after surgical interventions like vasectomy and catheterization surgery was also observed in other studies [39,40]. On the contrary, after nonsurgical interventions such as nicotine injection [32] or restrain stress [27,41,42] the corticosterone concentration usually declines within 60 to 120 min. The results indicate that the

long-lasting corticosterone response might be a characteristic feature for the stress response after surgical procedures.

Several reasons might account for these altered corticosterone kinetics. One reason could be the effect of anesthesia. Indeed, other studies reported a significant increase in corticosterone concentration within minutes of anesthesia using isoflurane. However, the increased corticosterone concentration was transient and declined after 60 min. Therefore, anesthesia alone is probably not responsible for the long-lasting activation of the HPA axis. Another reason might be that the summation of stressors such as pain, wound tension, tissue damage and the release of cytokines provokes a long lasting glucocorticoid response [43]. Moreover, the cytokine interleukin-6 (IL-6) can influence the HPA axis directly by stimulating the glucocorticoid release on the adrenal glands [43,44]. The release of IL-6 usually starts within 60 min after surgical procedures [45,46] and positively correlates with the corticosterone concentration [43,47]. In conclusion, the long-lasting corticosterone response after surgery might arise from the summation of distress parameters that influence the HPA axis.

One aim of this study was to evaluate which aspect can be used to judge animal distress: the maximal corticosterone concentration or rather the period of time for the significantly increased corticosterone response. Quantifying distress by our score sheet allowed us to rank the disease from highest distress (score 5) after BDL to lowest distress (0.75) during pancreatitis (Table 1). When comparing corticosterone concentration at 30 minutes between these animal models, a very similar rank order could be observed (Table 1). Indeed, the corticosterone concentrations of all animals, which were evaluated in all four animal models, correlated well with the distress score (Spearman's rank correlation coefficient: 0.67; $p = 0.003$). Previous studies support the hypothesis of a direct correlation between corticosterone concentration and stressor intensity [32,48,49]. Physical stress indicators, such as body weight loss, increased heart rate, and decreased food intake are also known to be associated with increased plasma corticosterone concentration. This emphasizes the relevance of this parameter for the analysis of distress [14,31,42]. However, since corticosterone is also influenced by other physical responses, i.e., circadian rhythm [3,8,9], estrus cycle [3,11,12], or sexual arousal [10], this parameter should always be assessed together with other stress or welfare parameters like distress score or body weight.

Furthermore, when comparing the duration of significantly increased corticosterone concentration between the four animal models a rank order could be established (Table 1). This rank order was identical to the rank order, when we scored distress (Table 1). This suggests that also the duration of increased corticosterone concentration is a relevant parameter for the analysis of distress. This conclusion is supported by various studies. For example, preventive wheel running shortened the glucocorticoid response after acute restrain stress and also reduced anxiety-like behavior [50]. Moreover, buprenorphine treatment decreases the duration as well as the corticosterone concentration after catheterization [40]. Therefore, we suggest that both parameters of corticosterone kinetics, absolute value of corticosterone concentration as well as duration of corticosterone response, are appropriate parameters for quantifying distress. Additionally, our findings indicate that corticosterone kinetics can be used as an important parameter for the grading of distress level between distinct animal models. Though the absolute value and duration of corticosterone kinetics correlate well with the assessed distress scoring, we want to mention that a direct comparison of two distinct mouse strains should be handled with care, as strain might act as a confounding variable. For example, BALB/c mice are known to show higher absolute corticosterone values and a longer duration compared to the C57BL/6J strain, which is probably caused by higher anxiety [5] or genetic alterations in the HPA axis [15]. Nevertheless, we got very similar results in both mouse strains after different interventions: Not only the absolute value of corticosterone concentration after 30 minutes but also the duration of the corticosterone response correlates well with the quantification of animal distress by a score sheet. We hope that the purposely increased heterogeneity in the experimental design improves external validity [51,52] and that these results will, therefore, provide an important basis to compare corticosterone concentrations of different animal models across mouse strains.

It is widely published that the time point of highest corticosterone concentration is dependent on the kind of intervention. Some studies noticed the peak of corticosterone response 20–30 min after the injection of toxic chemicals [16,32], or restrain stress [27,42]. However, after some other stressors like electric shocks [53] or acute shaker stress [14], the maximum of corticosterone concentration can be observed 60 min after distress. Thus, for comparing animal models, it is advisable to evaluate corticosterone kinetics previously on a few animals instead of relying on only one random time point. This procedure subsequently allows the reduction of animals. However, our data suggest that the corticosterone concentration at 30 minutes is a suitable parameter for comparing distress of our gastrointestinal disease models.

5. Conclusions

The present study demonstrated that distress caused by different interventions correlates with the peak and duration of corticosterone kinetics in distinct gastrointestinal mouse models. These results support the use of corticosterone kinetics as an important readout parameter to grade distress in animal models.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2615/9/4/145/s1, Table S1: Distress score on mice, Table S2: Consequences according to distress score.

Author Contributions: S.K., D.Z., B.V. designed the experiments. S.K., D.Z., G.T., X.Z., H.K., performed the experiments. S.K., D.Z., B.V. wrote the manuscript, and all authors approved the final version.

Funding: This study was supported by the Deutsche Forschungsgemeinschaft (DFG research group FOR 2591, project number: 321137804, ZE 712/1-1 and VO 450/15-1).

Acknowledgments: The authors are grateful for the perfect technical assistance from Berit Blendow (Institute for Experimental Surgery, University Medical Center, Rostock).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Haczeyni, F.; Yeh, M.M.; Ioannou, G.N.; Leclercq, I.A.; Goldin, R.; Dan, Y.Y.; Yu, J.; Teoh, N.C.; Farrell, G.C. Mouse Models of Nonalcoholic Steatohepatitis: A Reflection on Recent Literature. *J. Gastroenterol. Hepatol.* **2018**, *33*, 1312–1320. [CrossRef] [PubMed]
2. Zhan, X.; Wang, F.; Bi, Y.; Ji, B. Animal models of gastrointestinal and liver diseases. Animal models of acute and chronic pancreatitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2016**, *311*, G343–G355. [CrossRef]
3. Gong, S.; Miao, Y.-L.; Jiao, G.-Z.; Sun, M.-J.; Li, H.; Lin, J.; Luo, M.-J.; Tan, J.-H. Dynamics and correlation of serum cortisol and corticosterone under different physiological or stressful conditions in mice. *PLoS ONE* **2015**, *10*, e0117503. [CrossRef] [PubMed]
4. Munck, A.; Guyre, P.M.; Holbrook, N.J. Physiological functions of glucocorticoids in stress and their relation to pharmacological actions. *Endocr. Rev.* **1984**, *5*, 25–44. [CrossRef] [PubMed]
5. Benedetti, M.; Merino, R.; Kusuda, R.; Ravanelli, M.I.; Cadetti, F.; dos Santos, P.; Zanon, S.; Lucas, G. Plasma corticosterone levels in mouse models of pain. *Eur. J. Pain* **2012**, *16*, 803–815. [CrossRef]
6. De Kloet, E.R.; Vreugdenhil, E.; Oitzl, M.S.; Joëls, M. Brain corticosteroid receptor balance in health and disease. *Endocr. Rev.* **1998**, *19*, 269–301. [CrossRef]
7. Ingle, D.J. Permissibility of hormone action; a review. *Acta Endocrinol.* **1954**, *17*, 172–186.
8. Dalm, S.; Enthoven, L.; Meijer, O.C.; van der Mark, M.H.; Karssen, A.M.; de Kloet, E.R.; Oitzl, M.S. Age-related changes in hypothalamic-pituitary-adrenal axis activity of male C57BL/6J mice. *Neuroendocrinology* **2005**, *81*, 372–380. [CrossRef]
9. Teilmann, A.C.; Kalliokoski, O.; Sørensen, D.B.; Hau, J.; Abelson, K.S.P. Manual versus automated blood sampling: Impact of repeated blood sampling on stress parameters and behavior in male NMRI mice. *Lab. Anim.* **2014**, *48*, 278–291. [CrossRef]
10. Bronson, F.H.; Desjardins, C. Endocrine responses to sexual arousal in male mice. *Endocrinology* **1982**, *111*, 1286–1291. [CrossRef]

11. Nichols, D.J.; Chevins, P.F. Plasma corticosterone fluctuations during the oestrous cycle of the house mouse. *Experientia* **1981**, *37*, 319–320. [CrossRef]
12. Atkinson, H.C.; Waddell, B.J. Circadian variation in basal plasma corticosterone and adrenocorticotropin in the rat: Sexual dimorphism and changes across the estrous cycle. *Endocrinology* **1997**, *138*, 3842–3848. [CrossRef] [PubMed]
13. Bowers, S.L.; Bilbo, S.D.; Dhabhar, F.S.; Nelson, R.J. Stressor-specific alterations in corticosterone and immune responses in mice. *Brain Behav. Immun.* **2008**, *22*, 105–113. [CrossRef] [PubMed]
14. Bernatova, I.; Key, M.P.; Lucot, J.B.; Morris, M. Circadian differences in stress-induced pressor reactivity in mice. *Hypertension* **2002**, *40*, 768–773. [CrossRef]
15. Shanks, N.; Griffiths, J.; Zalcman, S.; Zacharko, R.M.; Anisman, H. Mouse strain differences in plasma corticosterone following uncontrollable footshock. *Pharmacol. Biochem. Behav.* **1990**, *36*, 515–519. [CrossRef]
16. Stanulis, E.D.; Matulka, R.A.; Jordan, S.D.; Rosecrans, J.A.; Holsapple, M.P. Role of corticosterone in the enhancement of the antibody response after acute cocaine administration. *J. Pharmacol. Exp. Ther.* **1997**, *280*, 284–291.
17. Williams, J.G.; Roberts, S.E.; Ali, M.F.; Cheung, W.Y.; Cohen, D.R.; Demery, G.; Edwards, A.; Greer, M.; Hellier, M.D.; Hutchings, H.A.; et al. Gastroenterology services in the UK. The burden of disease, and the organisation and delivery of services for gastrointestinal and liver disorders: A review of the evidence. *Gut* **2007**, *56* (Suppl. 1), 1–113. [CrossRef]
18. Majumder, S.; Chari, S.T. Chronic pancreatitis. *Lancet* **2016**, *387*, 1957–1966. [CrossRef]
19. Bataller, R.; Brenner, D.A. Hepatic stellate cells as a target for the treatment of liver fibrosis. *Semin. Liver Dis.* **2001**, *21*, 437–451. [CrossRef]
20. Pliarchopoulou, K.; Pectasides, D. Pancreatic cancer: Current and future treatment strategies. *Cancer Treat. Rev.* **2009**, *35*, 431–436. [CrossRef] [PubMed]
21. Zechner, D.; Bürtin, F.; Albert, A.-C.; Zhang, X.; Kumstel, S.; Schönrogge, M.; Graffunder, J.; Shih, H.-Y.; Müller, S.; Radecke, T.; et al. Intratumoral heterogeneity of the therapeutical response to gemcitabine and metformin. *Oncotarget* **2016**, *7*, 56395–56407. [CrossRef] [PubMed]
22. Bangen, J.-M.; Hammerich, L.; Sonntag, R.; Baues, M.; Haas, U.; Lambertz, D.; Longerich, T.; Lammers, T.; Tacke, F.; Trautwein, C.; et al. Targeting CCl4-induced liver fibrosis by RNA interference-mediated inhibition of cyclin E1 in mice. *Hepatology* **2017**, *66*, 1242–1257. [CrossRef] [PubMed]
23. Hashimoto, K.; Ethridge, R.T.; Saito, H.; Rajaraman, S.; Evers, B.M. The PPARgamma ligand, 15d-PGJ2, attenuates the severity of cerulein-induced acute pancreatitis. *Pancreas* **2003**, *27*, 58–66. [CrossRef] [PubMed]
24. Bleich, A.; Tolba, R.H. How can we assess their suffering? German research consortium aims at defining a severity assessment framework for laboratory animals. *Lab. Anim.* **2017**, *51*, 667. [CrossRef] [PubMed]
25. Zechner, D.; Knapp, N.; Bobrowski, A.; Radecke, T.; Genz, B.; Vollmar, B. Diabetes increases pancreatic fibrosis during chronic inflammation. *Exp. Biol. Med.* **2014**, *239*, 670–676. [CrossRef] [PubMed]
26. Abshagen, K.; Rotberg, T.; Genz, B.; Vollmar, B. No significant impact of Foxf1 siRNA treatment in acute and chronic CCl4liver injury. *Exp. Biol. Med.* **2017**, *242*, 1389–1397. [CrossRef] [PubMed]
27. Vahl, T.P.; Ulrich-Lai, Y.M.; Ostrander, M.M.; Dolgas, C.M.; Elfers, E.E.; Seeley, R.J.; D'Alessio, D.A.; Herman, J.P. Comparative analysis of ACTH and corticosterone sampling methods in rats. *Am. J. Physiol. Endocrinol. Metab.* **2005**, *289*, E823–E828. [CrossRef] [PubMed]
28. Kugler, J.; Lange, K.W.; Kalveram, K.T. Influence of bleeding order on plasma corticosterone concentration in the mouse. *Exp. Clin. Endocrinol.* **1988**, *91*, 241–243. [CrossRef]
29. Paster, E.V.; Villines, K.A.; Hickman, D.L. Endpoints for mouse abdominal tumor models: Refinement of current criteria. *Comp. Med.* **2009**, *59*, 234–241. [PubMed]
30. Razali, N.M. Power comparison of Shapri-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Stat. Model. Anal.* **2011**, 21–33.
31. Taylor, B.K.; Akana, S.F.; Peterson, M.A.; Dallman, M.F.; Basbaum, A.I. Pituitary-adrenocortical responses to persistent noxious stimuli in the awake rat: Endogenous corticosterone does not reduce nociception in the formalin test. *Endocrinology* **1998**, *139*, 2407–2413. [CrossRef] [PubMed]
32. Freund, R.K.; Martin, B.J.; Jungschaffer, D.A.; Ullmann, E.A.; Collins, A.C. Genetic Differences in Plasma Corticosterone Levels in Response to Nicotine injection. *Pharmacol. Biochem. Behav.* **1988**, *1988*, 1059–1064. [CrossRef]

33. Herman, J.P.; McKlveen, J.M.; Solomon, M.B.; Carvalho-Netto, E.; Myers, B. Neural regulation of the stress response: Glucocorticoid feedback mechanisms. *Braz. J. Med Biol. Res. = Rev. Bras. Pesqui. Med. E Biol.* **2012**, *45*, 292–298. [CrossRef]
34. Di, S.; Malcher-Lopes, R.; Halmos, K.C.; Tasker, J.G. Nongenomic glucocorticoid inhibition via endocannabinoid release in the hypothalamus: A fast feedback mechanism. *J. Neurosci.* **2003**, *23*, 4850–4857. [CrossRef] [PubMed]
35. Keller-Wood, M.E.; Dallman, M.F. Corticosteroid inhibition of ACTH secretion. *Endocr. Rev.* **1984**, *5*, 1–24. [CrossRef]
36. Kim, D.H.; Jung, J.S.; Song, D.K.; Suh, H.W.; Huh, S.O.; Kim, Y.H. Intracerebroventricular injection-induced increase in plasma corticosterone levels in the mouse: A stress model. *J. Pharmacol. Toxicol. Methods* **1998**, *39*, 71–73. [CrossRef]
37. Allen, J.P.; Allen, C.F. The Effect of Dimethyl Sulfoxide on Hypothalamic-Pituitary-Adrenal Fuctions in Rat. *Ann. N. Y. Acad. Sci.* **1975**, *243*, 325–336. [CrossRef]
38. Ivanović Matić, S.; Dinić, S.; Mihailović, M.; Grigorov, I.; Bogojević, D.; Poznanović, G. Acute-phase protein expression in DMSO-intoxicated rats. *Toxicol. Lett.* **2004**, *147*, 153–159. [CrossRef] [PubMed]
39. Jacobsen, K.R.; Kalliokoski, O.; Teilmann, A.C.; Hau, J.; Abelson, K.S.P. The effect of isoflurane anaesthesia and vasectomy on circulating corticosterone and ACTH in BALB/c mice. *Gen. Comp. Endocrinol.* **2012**, *179*, 406–413. [CrossRef]
40. Sundbom, R.; Jacobsen, K.R.; Kalliokoski, O.; Hau, J.; Abelson, K.S.P. Post-operative corticosterone levels in plasma and feces of mice subjected to permanent catheterization and automated blood sampling. *In Vivo* **2011**, *25*, 335–342. [PubMed]
41. Solomon, M.B.; Loftspring, M.; de Kloet, A.D.; Ghosal, S.; Jankord, R.; Flak, J.N.; Wulsin, A.C.; Krause, E.G.; Zhang, R.; Rice, T.; et al. Neuroendocrine Function After Hypothalamic Depletion of Glucocorticoid Receptors in Male and Female Mice. *Endocrinology* **2015**, *156*, 2843–2853. [CrossRef]
42. Fediuc, S.; Campbell, J.E.; Riddell, M.C. Effect of voluntary wheel running on circadian corticosterone release and on HPA axis responsiveness to restraint stress in Sprague-Dawley rats. *J. Appl. Physiol.* **2006**, *100*, 1867–1875. [CrossRef]
43. Desborough, J.P. The stress response to trauma and surgery. *Br. J. Anaesth.* **2000**, *85*, 109–117. [CrossRef] [PubMed]
44. Salas, M.A.; Evans, S.W.; Levell, M.J.; Whicher, J.T. Interleukin-6 and ACTH act synergistically to stimulate the release of corticosterone from adrenal gland cells. *Clin. Exp. Immunol.* **1990**, *79*, 470–473. [CrossRef] [PubMed]
45. Berger, K.; Sander, M.; Spies, C.D.; Weymann, L.; Bühner, S.; Lochs, H.; Wernecke, K.-D.; von Heymann, C. Profound haemodilution during normothermic cardiopulmonary bypass influences neither gastrointestinal permeability nor cytokine release in coronary artery bypass graft surgery. *Br. J. Anaesth.* **2009**, *103*, 511–517. [CrossRef]
46. Swartbol, P.; Pärsson, H.; Truedsson, L.; Sjöholm, A.; Norgren, L. Aortobifemoral surgery induces complement activation and release of interleukin-6 but not tumour necrosis factor-alpha. *Cardiovasc. Surg.* **1996**, *4*, 483–491. [CrossRef]
47. Leon, L.R.; Blaha, M.D.; DuBose, D.A. Time course of cytokine, corticosterone, and tissue injury responses in mice during heat strain recovery. *J. Appl. Physiol.* **2006**, *100*, 1400–1409. [CrossRef] [PubMed]
48. Armario, A.; Montero, J.L.; Balasch, J. Sensitivity of corticosterone and some metabolic variables to graded levels of low intensity stresses in adult male rats. *Physiol. Behav.* **1986**, *37*, 559–561. [CrossRef]
49. Kant, G.J.; Leu, J.R.; Anderson, S.M.; Mougey, E.H. Effects of chronic stress on plasma corticosterone, ACTH and prolactin. *Physiol. Behav.* **1987**, *40*, 775–779. [CrossRef]
50. Hare, B.D.; Beierle, J.A.; Toufexis, D.J.; Hammack, S.E.; Falls, W.A. Exercise-associated changes in the corticosterone response to acute restraint stress: Evidence for increased adrenal sensitivity and reduced corticosterone response duration. *Neuropsychopharmacology* **2014**, *39*, 1262–1269. [CrossRef]
51. Voelkl, B.; Vogt, L.; Sena, E.S.; Würbel, H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLoS Biol.* **2018**, *16*, e2003693. [CrossRef] [PubMed]

52. Richter, S.H.; Garner, J.P.; Würbel, H. Environmental standardization: Cure or cause of poor reproducibility in animal experiments? *Nat. Methods* **2009**, *6*, 257–261. [CrossRef]
53. Dominguez, G.; Faucher, P.; Henkous, N.; Krazem, A.; Piérard, C.; Béracochéa, D. Stress induced a shift from dorsal hippocampus to prefrontal cortex dependent memory retrieval: Role of regional corticosterone. *Front. Behav. Neurosci.* **2014**, *8*, 166. [CrossRef]

scientific reports

Check for updates

OPEN

Comparing distress of mouse models for liver damage

Guanglin Tang[1], Nico Seume[1], Christine Häger[2], Simone Kumstel[1], Kerstin Abshagen[1], André Bleich[2], Brigitte Vollmar[1], Steven R. Talbot[2,3]✉, Xianbin Zhang[1,3]✉ & Dietmar Zechner[1,3]✉

In order to foster animal welfare as well as high quality of research, many countries regulate by law that the severity of animal experiments must be evaluated and considered when performing biomedical research. It is well accepted that multiple parameters rather than a single readout parameter should be applied to describe animal distress or suffering. However, since the performance of readout parameters for animal distress is rarely defined and methods for multivariate analysis have only in rare cases been used, it is not known which methodology is most appropriate to define animal distress. This study used receiver operating characteristic curve analysis to quantify the performance of burrowing activity, body weight change and a distress score of mice after induction of liver damage by bile duct ligation or carbon tetrachloride. In addition, Support Vector Machine classification was used to compare the distress of these mouse models. This approach demonstrated that bile duct ligation causes much more distress than carbon tetrachloride-induced liver damage. This study, therefore, provides a prototype how to compare two animal models by considering several readout parameters. In the future these or similar methods for multivariate analysis will be necessary, when assessing and comparing the severity of animal models.

Public discussions on animal welfare have caused the implementation of laws and guidelines to regulate experiments on animals in most countries[1,2]. This made animal welfare a top priority when conducting and publishing in vivo studies[3–5]. Thus, when pursuing animal experiments, scientists have to balance two goals: animal welfare and the potential benefit of research. While this objective is self-evident and coherent, a detailed concept what needs to be done to balance both goals is more difficult to define. In many countries a prospective and often also actual severity assessment of animal experiments are legally required[6]. This should provide the basis for an ethical evaluation and the conclusion, if an animal experiment is justified and, therefore, should be allowed to be conducted.

Thus, an evidence-based analysis of animal distress is often legally required and is also essential for a realistic harm/benefit analysis, a sensible selection of an animal model and the development of refinement strategies. Scientists have primarily used non-invasive methods to assess animal distress. For example, many distress scores based on appearance, behaviour and physical parameters of rodents have been developed[7–9]. In addition, natural behaviour of animals such as burrowing activity has been explored to assess distress[10–12]. One of the most popular parameters to evaluate suffering from animals is body weight which has the distinct advantage that it can be easily and objectively measured[7,13–15].

While many distinct readout parameters for measuring distress are available, very little is known about how these methods can be compared. The performance of a method or a diagnostic test is usually evaluated by receiver operating characteristic (ROC) curve analysis. The area under the curve (AUC) quantifies this performance and indicates how accurately a test discriminates between two states, typically referred to as diseased and non-diseased state[16]. However, it is well accepted that multiple parameters rather than a single readout parameter should be applied to describe and compare animal distress[7,17,18]. Many studies indeed evaluate several readout parameters for distress, but do not combine these parameters by a statistical procedure to reach a holistic conclusion[13,19–22]. To facilitate such an integrated conclusion, a multivariate analysis, which combines different readout parameters when analysing animal distress, is necessary. Such analyses are often performed in clinical situations in form of a binary logistic regression in order to test whether a combination of biomarkers has higher discriminatory power to differentiate between diseased and non-diseased states than single biomarkers[23,24]. Another option to analyse

[1]Rudolf-Zenker, Institute of Experimental Surgery, Rostock University Medical Center, Rostock, Germany. [2]Institute for Laboratory Animal Science, Hannover Medical School, Hanover, Germany. [3]These authors jointly supervised this work: Steven R. Talbot, Xianbin Zhang and Dietmar Zechner. ✉email: talbot.steven@mh-hannover.de; zhangxianbin@hotmail.com; dietmar.zechner@uni-rostock.de

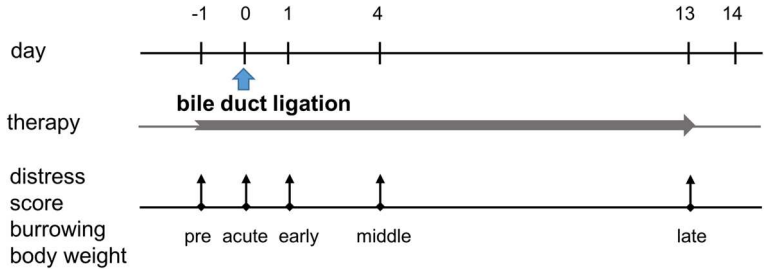


Figure 1. Scheme describing how the experiments for the bile duct ligation (BDL) animal model were performed. Surgery was done on day 0 and the distress parameters were evaluated during pre, acute, early, middle and late phase on the indicated days. Therapies with MCC950 or vehicle solution were performed by daily subcutaneous injection from day − 1 to day 13.

more than one readout parameter simultaneously is clustering, followed by Support Vector Machine (SVM) classification. For example, clustering was used to differentiate between subgroups of patients with irritable bowel syndrome[25] or to compare distinct distress levels of mice during colitis[15].

Thus, it was one aim of this study to evaluate, if ROC curve analysis and binary logistic regression be used to describe the performance of single or multiple readout parameters for defining distress in animals. Moreover, it was the aim to assess whether SVM classification can be used to compare the severity of two animal models. We compared distress caused by bile duct ligation (BDL) to distress caused by carbon tetrachloride (CCl_4). These two animal models are widely used for studying liver damage and fibrosis[26–30].

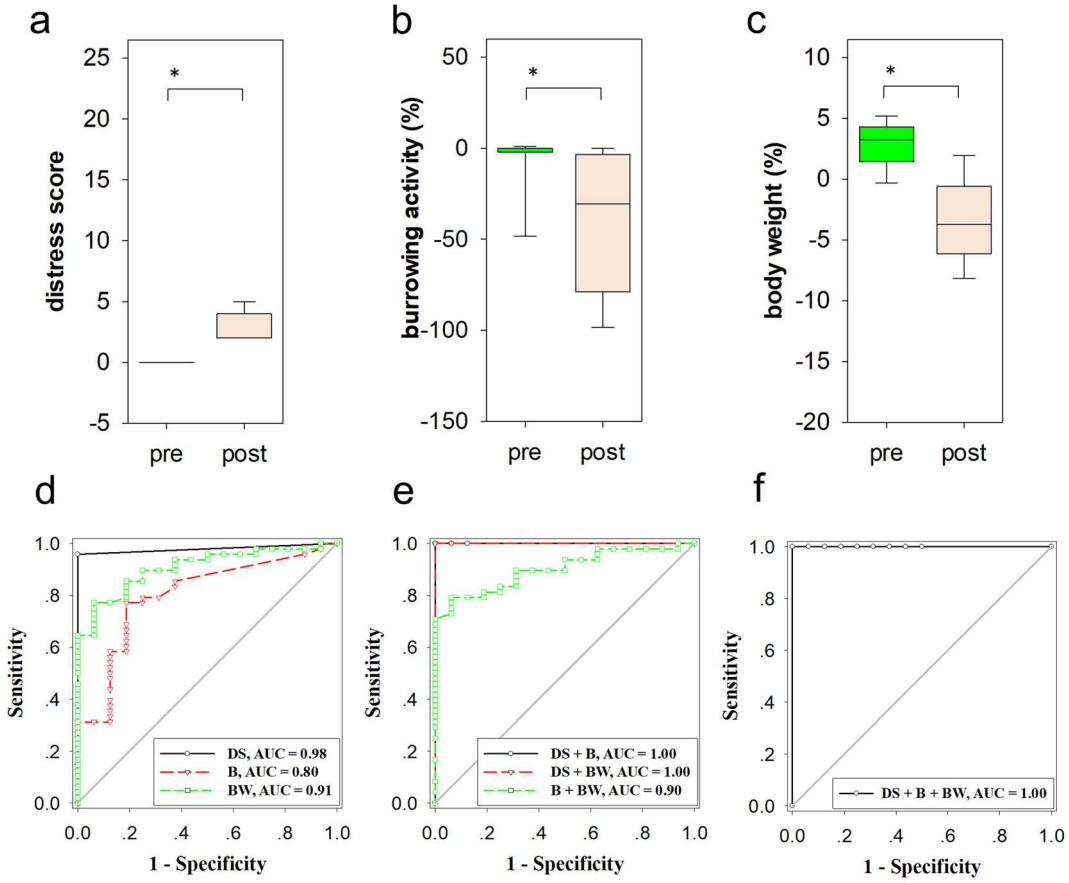
Results

Characterisation of parameters measuring distress after BDL. Mice were evaluated before and after BDL during the early, middle and late phases of cholestasis by assessing a distress score, burrowing activity and body weight (Fig. 1). First of all, we aspired to evaluate the suitability of these parameters to measure distress of mice. We hypothesized that parameters, which are suitable to measure distress should be able to differentiate between healthy and diseased mice as well as between mice which survived and non-survivors.

Thus, we first analysed mice, which survived until day 14 (survivors), in order to explore, if these read out parameters could differentiate between healthy and diseased mice. While the distress score increased continuously after BDL, the burrowing activity and body weight of mice rather decreased after this intervention (supplementary Fig. S1). No significant change in any of these parameters was observed when treating the mice with the NLRP3 inflammasome inhibitor MCC950 (supplementary Fig. S1), although previous studies suggested that this inhibitor can have analgesic function[31]. Thus, all BDL cohort mice were pooled and distress before BDL (pre) was compared to distress after BDL (post). We observed that BDL led to a significant increase of the distress score (Fig. 2a). It caused a significant decrease of burrowing activity (Fig. 2b) and a reduction of body weight (Fig. 2c). This suggests that distress score, burrowing activity and change in body weight are sensitive parameters that can differentiate between distress before (level 0) and after BDL (level 1). To evaluate the performance of these parameters in distinguishing between these two distress levels, we used ROC curves. We observed that all parameters, distress score, burrowing activity and body weight, can discriminate between these two distress levels (Fig. 2d). Combining multiple distress parameters with binary logistic regression revealed that the combination of distress score plus burrowing activity, distress score plus body weight and the combination of all three parameters produced a very high AUC indicating a very good performance in defining distress (Fig. 2e–g).

We also evaluated, if distress parameters could differentiate between different magnitudes of cholestasis. ALP activity has been demonstrated to increase with the progression of cholestasis[32]. Therefore, we evaluated ALP activity of mice after 2, 5 or 14 days of cholestasis and used k-means clustering to discretize the data into two categories: Low ALP and high ALP. Surprisingly, we observed that neither the distress score nor the burrowing activity could differentiate between low ALP and high ALP animals (supplementary Fig. S2). However, body weight change could differentiate well (AUC = 0.79) between these two clusters (supplementary Fig. S2). In order to analyse, if other parameters measuring distress would improve the differentiation between low ALP and high ALP animals, we determined the corticosterone concentration in the blood plasma (supplementary Fig. S2). Indeed, the corticosterone concentration in the blood plasma of animals could also differentiate well (AUC = 0.72) between low and high ALP animals (supplementary Fig. S2). However, when combining body weight change and corticosterone concentration in a logistic regression the discriminatory power of the combination was not higher than the discriminatory power of only the body weight change (supplementary Fig. S2). Thus, for differentiating between low and high ALP animals analysing body weight change is sufficient. Possibly, a combination with yet unknown additional distress parameters might be needed to predict the magnitude of cholestasis with an even higher discriminatory power.

We then explored, if mice which did not survive until day 14 (non-survivors) reached a different distress level before death when compared to mice that survived after BDL. We observed that the distress score of non-survivors measured before death is significantly higher than the distress score of survivors (Fig. 3a). The burrowing activity (Fig. 3b) and body weight (Fig. 3c) of non-survivors were significantly lower than those of surviving mice. These data suggest that non-survivors experience increased distress before death (level 2) when compared to surviving mice (level 1). In order to evaluate the performance of the readout parameters in

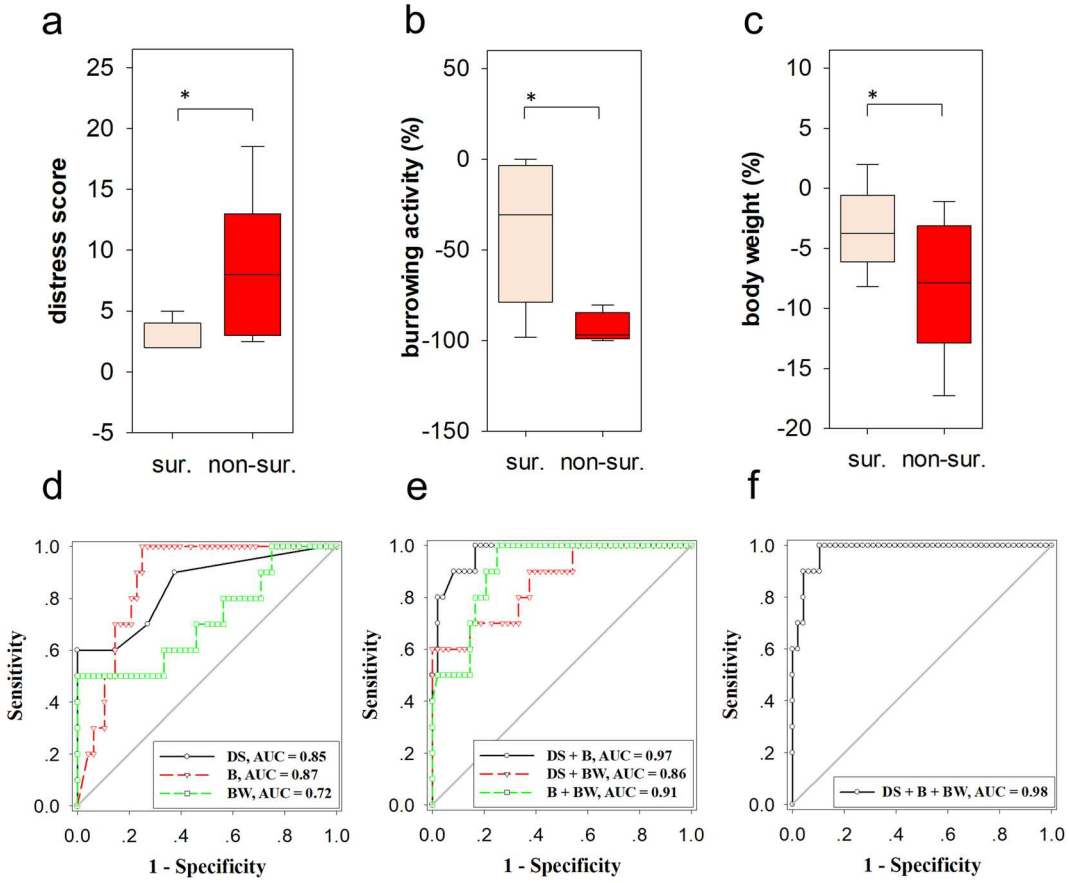


g

	Variables	AUC	95 % CI	Asymptotic *P*-value	Number of data points
Single	Distress score	0.979	0.951 - 1.000	< 0.0001	pre = 16 post = 48
	Burrowing activity (%)	0.800	0.676 - 0.925	0.0004	pre = 16 post = 48
	Body weight (%)	0.907	0.834 - 0.980	< 0.0001	pre = 16 post = 48
Double	Distress score + Burrowing activity (%)	1.000	1.000 - 1.000	< 0.0001	pre = 16 post = 48
	Distress score + Body weight (%)	1.000	1.000 - 1.000	< 0.0001	pre = 16 post = 48
	Burrowing activity (%) + Body weight (%)	0.900	0.827 - 0.974	< 0.0001	pre = 16 post = 48
Triple	Distress score + Burrowing activity (%) + Body weight (%)	1.000	1.000 - 1.000	< 0.0001	pre = 16 post = 48

Figure 2. Distress before and after BDL. The distress score (**a**) was increased (Mann–Whitney rank sum test, $P \leq 0.001$), burrowing activity (**b**) was decreased (Mann–Whitney rank sum test, $P \leq 0.001$) and body weight (**c**) was also decreased (Mann–Whitney rank sum test, $P \leq 0.001$), when comparing data taken before BDL (pre) to data taken after BDL (post). ROC curve analysis that computed the area under the curve (AUC) for single (**d**), two (**e**) or all three (**f**) distress parameters. The performance of single and multiple parameters is described by presenting the AUC, the 95% confidence interval (CI) and the asymptotic P-value (**g**). Data of 16 mice, pre: n = 16 data points, post: n = 48 data points.

distinguishing between these two distress levels, we used ROC curves. All single readout parameters such as distress score, burrowing activity and change in body weight had discriminatory power to differentiate between survivors and non-survivors (Fig. 3d). After combining multiple distress parameters with binary logistic regression, we observed that combination of two or three parameters also had a high discriminatory power (Fig. 3e,f).

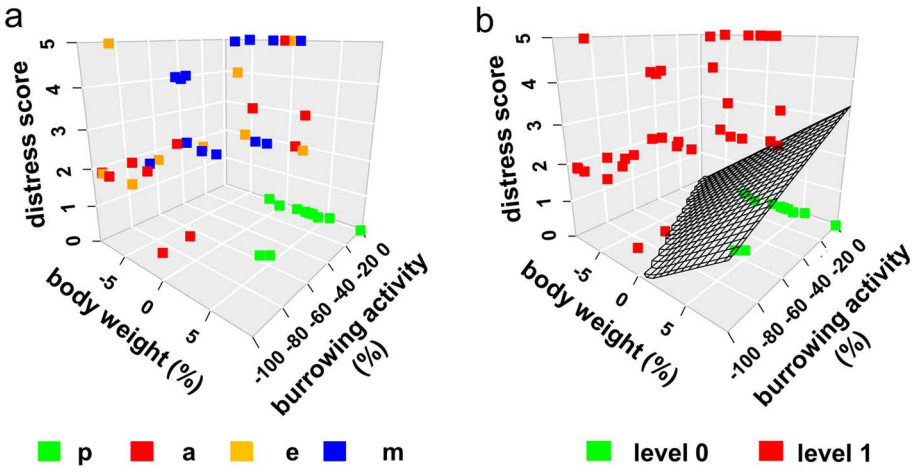


g

	Variables	AUC	95 % CI	Asymptotic *P*-value	Number of data points
Single	Distress score	0.848	0.701 -0.995	0.0006	sur. = 48 non-sur. = 10
	Burrowing activity (%)	0.871	0.781 - 0.961	0.0002	sur. = 48 non-sur. = 10
	Body weight (%)	0.719	0.514 - 0.924	0.0307	sur. = 48 non-sur. = 10
Double	Distress score + Burrowing activity (%)	0.971	0.931 - 1.000	< 0.0001	sur. = 48 non-sur. = 10
	Distress score + Body weight (%)	0.860	0.727 - 0.994	0.0004	sur. = 48 non-sur. = 10
	Burrowing activity (%) + Body weight (%)	0.907	0.826 - 0.988	< 0.0001	sur. = 48 non-sur. = 10
Triple	Distress score + Burrowing activity (%) + Body weight (%)	0.979	0.948 - 1.000	< 0.0001	sur. = 48 non-sur. = 10

Figure 3. Distress of survivors and non-survivors after BDL. The distress score (**a**) was increased (Mann–Whitney rank sum test, $P \leq 0.001$), whereas burrowing activity (**b**) was decreased (Mann–Whitney rank sum test, $P \leq 0.001$) and body weight (**c**) was also reduced (Mann–Whitney rank sum test, $P = 0.031$), when comparing data of survivors (sur.) to data of non-survivors (non-sur.). ROC curve analysis shows the area under the curve (AUC) for single (**d**), two (**e**) or all three (**f**) distress parameters. The performance of single and multiple parameters is described by presenting the AUC, the 95% confidence interval (CI) and the asymptotic P-value (**g**). Survivors: 16 mice, 48 data points; non-survivors: 10 mice, 10 data points.

The combination of all three parameters (distress score plus burrowing activity plus body weight) produced the largest AUC, suggesting that the combination of all readout parameters allows the best differentiation between survivors and non-survivors (Fig. 3g). These data, therefore, suggest that the distress score, burrowing activity and body weight are suitable parameters to describe distinct distress levels.



c

	body weight (%) + burrowing activity (%) + distress score			
	not optimized training data	not optimized test data	optimized training data	optimized test data
accuracy	1	1	1	0.85
sensitivity	1	1	1	1
specificity	1	1	1	0.80

Figure 4. Generation of a training model by SVM. Single data points (squares), which were derived from the training data set from identical animal previous (p) to BDL and at the acute (a), early (e), and middle (m) phase of cholestasis are presented in form of a three dimensional scatter plot (**a**). A discriminatory model was built by training a linear SVM kernel to the labelled data in order to differentiate between two levels (level 0 and level 1) of distress (**b**): The resulting classifier (hyperplane) discriminates between these two levels. The accuracy, sensitivity and specificity of the training model was characterised using either the training data themselves or a test data set and applying the hyperplane (not optimized) or an optimized hyperplane after a 5-times repeated tenfold cross validation (**c**). Training data set: n = 11 data points (pre), post: n = 33 data points (post); test data set: n = 5 data points (pre), post: n = 15 data points (post).

Considering multiple parameters when differentiating between two distress levels. Next, we evaluated whether all three parameters can be used together to discriminate between the distress of healthy (pre-intervention) against the distress of diseased animals (post-intervention). We used machine learning to address this question: more specifically, we used a Support Vector Machine (SVM) to classify samples. Class-labels were obtained by labelling pre- against post-intervention data. For subsequent classification, we first split the data randomly into a training (containing 70% of data) and a test data set (containing 30% of data). The model was then built using the training data (Fig. 4a). Within the SVM, a linear kernel function was used to find the classifier. This tuned and optimized discriminator was visualized in the plots as a hyperplane, separating two putative levels of distress, which were defined as distress level 0 or distress level 1 (Fig. 4b).

For internal model optimization, and to address potential sampling bias we used hyper-parameter tuning and fivefold repeated tenfold cross-validation. The mean accuracies, sensitivities and specificities from this process were reported for the model (Fig. 4c shows results for both, the optimized and non-optimized model). The model itself was validated using the excluded (and labelled) test data (Fig. 4c). We observed high accuracy, sensitivity, and specificity for training as well as test data (Fig. 4c). This suggests that the combination of all three parameters (distress score, burrowing activity, bodyweight) exhibits a high diagnostic ability for the differentiation between distress level 0 and distress level 1. The rigorous model design and cross-validation process further ensured that these results are not based on potential sampling bias. Also, the optimized model shows lower accuracies for the external test data (accuracy optimized model: 0.80; accuracy not optimized model: 1). This was expected as the not-optimized models tend to overfit the data.

Comparing distress of the BDL to the CCl_4 animal model. Next we pursued the question if and how we can compare the distress between two animal models. In order to compare the BDL model to another animal model widely used for studying liver damage and fibrosis, mice were repetitively injected with CCl_4 (Fig. 5a). These mice were also either treated with MCC950 or a vehicle control and the distress of these animals was analysed before any intervention and during the early, middle and late phases of disease progression by assessing the distress score, burrowing activity and body weight (Fig. 5a). Again, no significant change in distress score, burrowing activity and body weight was observed when treating the mice with MCC950 or a vehicle control (data not shown). Thus, all CCl_4 cohort mice were pooled and post-CCl_4 and post-BDL data were then compared

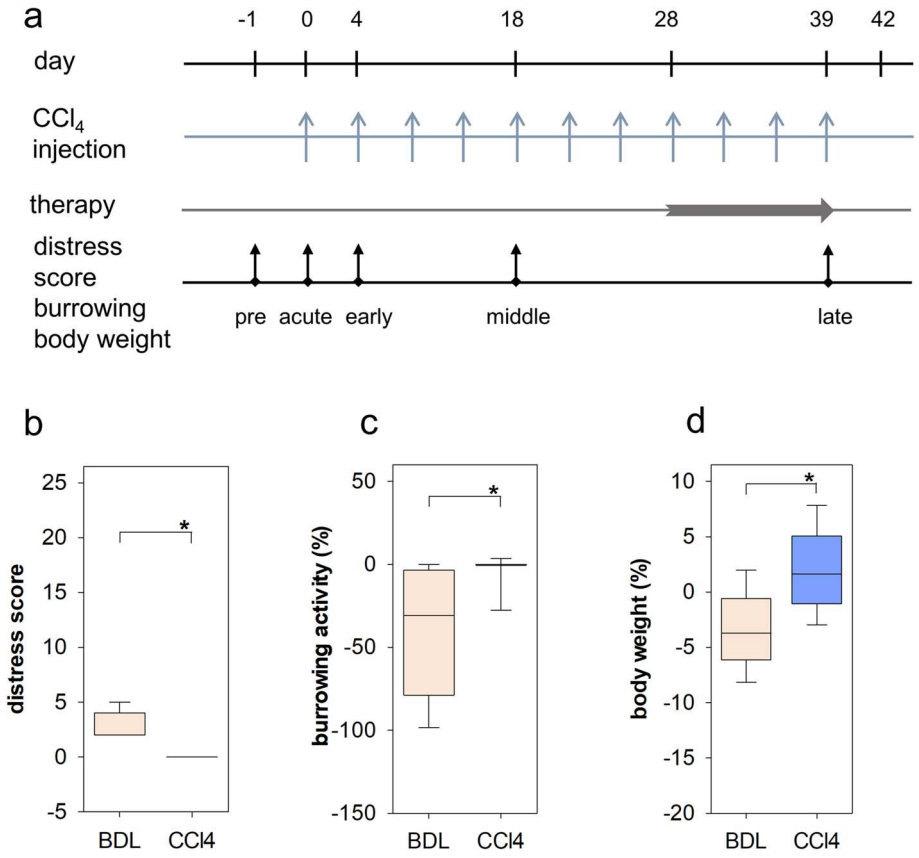


Figure 5. Distress of mice after CCl_4 injections and after bile duct ligation. Scheme describing how the experiments for the CCl_4 animal model were performed (**a**): CCl_4 was injected sc on the indicated days, the distress parameters were evaluated during pre, acute, early, middle and late phase and a therapy with MCC950 or vehicle solution was performed as daily subcutaneous injections from day 28 to 39. The distress score (**b**) was decreased (Mann–Whitney rank sum test, $P \leq 0.001$), whereas burrowing activity (**c**) was increased (Mann–Whitney rank sum test, $P \leq 0.001$) and body weight loss (**d**) was decreased (Student's t test, $P \leq 0.001$) when comparing post-BDL to post-CCl_4 animals. BDL: 16 mice = 48 data points; CCl_4: 10 mice = 30 data points.

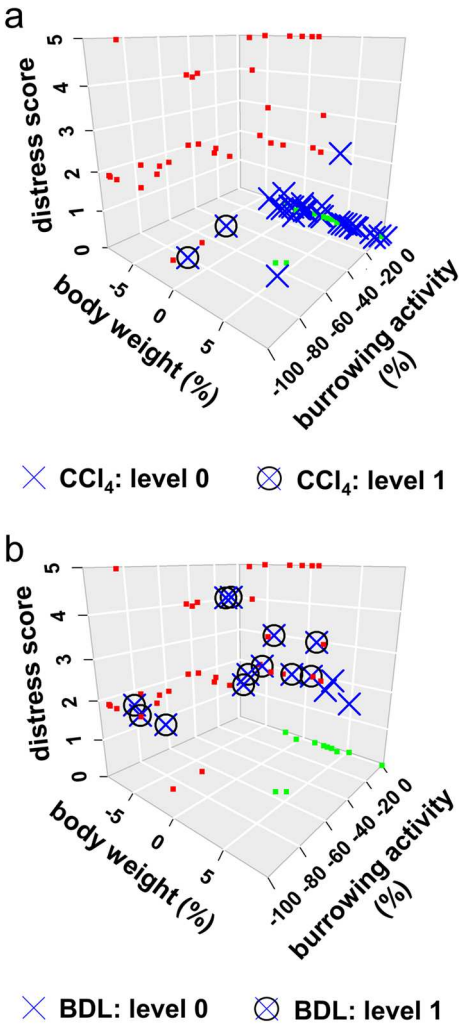
(Fig. 5b–d). We observed that CCl_4-treated mice had a significantly decreased distress score (Fig. 5b), increased burrowing activity (Fig. 5c) and significantly less body weight reduction (Fig. 5d), when compared to BDL mice. Thus, all three read out parameters indicate that CCl_4 causes less distress than BDL.

We then compared these two animal models by using the optimized training model based on 70% of the BDL data. We then classified the post-CCl_4 data according to this training model (see blue crosses in Fig. 6a). In addition, we classified the post-BDL data of the test data set (see blue crosses in Fig. 6b). Only 2 out of 30 post-CCl_4 data points were assigned to distress level 1, whereas 12 out of 15 post-BDL data points were correctly assigned to distress level 1 (Fig. 6c). Using Fisher's exact test, a significant difference in the distress levels distribution between BDL and the CCl_4 cohort was observed ($P < 0.001$). This multivariate analysis suggests that at most time points CCl_4-treated animals experience less distress than animals after BDL.

In order to compare liver damage in both animal models, we assessed the activity of aspartate aminotransferase (AST), alanine aminotransferase (ALT) and glutamate dehydrogenase (GLDH) in blood plasma. AST and ALT activity was significantly increased in cholestatic as well as CCl_4-treated mice, when compared to heathy control animals (supplementary Figure S3). GLDH was significantly increased in cholestatic animals when compared to healthy or CCl_4-treated mice (supplementary Figure S3). In addition, we also evaluated oxidative stress by measuring malondialdehyde in liver tissue. Malondialdehyde was significantly increased after repetitive CCl_4-treatment when compared to cholestatic or healthy mice (supplementary Figure S3). These results demonstrate that the liver is damaged after cholestasis and toxic liver injury, but that specific pathophysiological features such as the induction of oxidative stress differs between these two animal models.

Discussion

There is an urgent need to evaluate the feasibility of methods to compare distress caused by different animal models[33]. The present study compared BDL to CCl_4-induced liver damage and evaluated animal distress based on three distinct readout parameters. The multivariate analysis using SVM clearly demonstrated that BDL caused more distress than the treatment with CCl_4.



c

	CCl_4	BDL
level 0	28	3
level 1	2	12

*significantly different distribution of distress levels between CCl_4 and BDL

Figure 6. Comparing distress caused by CCl_4 injections or bile duct ligation using SVM classification. In both plots green and red squares indicate distress level 0 or distress level 1 of the BDL training data set and crosses denote data classified as distress level 0, whereas circled crosses denote data classified as distress level 1. Blue crosses denote post-CCl_4 (**a**) and post-BDL (**b**) distress. A 2 × 2 contingency table compares the distributions of predicted distress levels of post-CCl_4 to the post-BDL test data set (**c**). A significantly different distribution of distress levels between these data sets has been determined by Fisher's Exact Test, *$P \leq 0.001$), CCl_4: n = 30 data points; BDL: n = 15 data points.

No direct multivariate comparison of distress between BDL and CCl_4-induced liver damage has been published to our knowledge. However, publications describe an average body weight loss of 15–20%, 18%, or 20–30% after BDL[34–36] or a transient body weight loss of approximately 8% or 10% during repetitive CCl_4 injection[37,38]. This supports our conclusion that BDL causes more distress than CCl_4. However, the BDL animal model will still be needed for the following reasons. Distinct animal models are necessary to address the central principle of science that robust research needs many independent lines of evidence[39]. Indeed, BDL and CCl_4-induced liver damage are often used in one study to prove a scientific conclusion in two independent animal models[40,41]. In addition, there are also some differences between these two animal models. BDL causes an increase in biliary pressure, inflammation and cytokine secretion resulting in proliferation of biliary epithelial cells and portal fibrosis[42]. BDL therefore mimics cholestatic injury, which is, for example, observed during autoimmune diseases (primary biliary cirrhosis and primary sclerosing cirrhosis) and obstructive conditions such as cholelithiasis and

tumour compression of bile ducts[43]. In contrast, metabolites of CCl_4, such as trichloromethyl radicals, induce oxidative stress, centrilobular liver necrosis, an inflammatory response and liver fibrosis[42,44]. In many aspects, it mimics liver damage in humans by different toxins[42]. These distinct pathophysiological features and mechanisms of animal models will remain to be of utterly importance, when deciding which animal model will be used for addressing a specific scientific hypothesis. However, at least for the BDL animal model, the use of analgesics should be essential[4]. It is especially necessary to mention this point, if one considers that only 3.4% of studies, which describe experiments using BDL in mice, specified the administration of a systemic analgesic[45]. This is surprising, considering that it was already demonstrated decades ago that animals experience post-operative pain after BDL[46]. However, analgesia can also interfere with disease mechanisms and can actually be harmful to animals when applied in high doses[47,48].

The most important prerequisite for being able to judge animal distress are methods with high discriminatory power to differentiate between distinct distress levels. This study used ROC curve analysis to evaluate the discriminatory power of readout parameters. This tool has been widely used to define the diagnostic ability of methods in a clinical situation. For example, ROC curve analysis helped to define which biomarker in the blood has the best discriminatory power to predict pancreatic cancer[23] or which biochemical marker is suitable to predict increased risk of stillbirth in women with intrahepatic cholestasis of pregnancy[49]. In our study ROC curve analysis judged the suitability of readout parameters to differentiate between healthy mice and diseased mice or between diseased mice, which survive, and diseased mice, which will succumb to their disease. All readout parameters: distress score, burrowing activity and body weight change had discriminatory power to differentiate between animals before and after induction of cholestasis (Fig. 2d). However, burrowing activity was the parameter with the lowest performance (performance of parameters: distress score > body weight change > burrowing activity). When differentiating between survivors and non-survivors all readout parameters had again a high discriminatory power (Fig. 3d), but body weight change was the parameter with the lowest performance (performance of parameters: burrowing activity > distress score > body weight change). In addition to assessing the discriminatory power, one can determine the optimal cut-off of a diagnostic method by Youden's index and calculate the positive predictive value (PPV)[50]. We, therefore, also calculated the PPV using the combination of all three parameters. An optimal cut-off calculated by Youden's index lead to 5 false positive and 10 true positive predictions, resulting in a PPV of 67%. Thus, it is not practical to use this method for deciding, if animals should be euthanized, because one would kill too many animals, which would otherwise survive. However, the combination of all three parameters is useful in describing distinct distress levels and can be used to compare 2 different animal models. These experiments also demonstrate that not a single readout parameter can be used as the gold standard for all situations.

This need for considering multiple parameters to assess animal welfare was often postulated[7,17,18]. However, in many studies several parameters are evaluated, but these parameters are often not combined by a statistical procedure to reach a holistic conclusion[13,19–22]. Only very few studies exist, which use biostatistical methods to combine distinct readout parameters for defining animal distress. For example, Peng et al. have used composite z scores to compare the results of several behavioural tests between control mice and mice after surgery[20]. Häger et al. have used k-means clustering to compare distinct distress levels during colitis[15]. Möller et al. have used principal component analysis to describe many behavioural and biochemical variables supporting the conclusion that there is no major difference in distress between rats after electrode implantation and rats after electrode implantation plus kindling of epilepsy[51]. In our study we plotted three parameters and defined distress levels by SVM classification. This method had a high specificity, sensitivity and accuracy when validated with test data (Fig. 4c). However, we also want to emphasize that ROC curve analysis indicated that single read out parameters or two read out parameters, which were combined by multiple logistic regression, have also a very high discriminatory power to differentiate between distress levels in the BDL animal model (Fig. 2g). This indicates that less than three readout parameters might suffice to define the distress of animals and to compare animal models. However, we propose that substantiating a conclusion by considering several readout parameter is better to than relying on only one single parameter. Such a multivariate conclusion reduces arbitrariness when choosing a readout parameter and therefore diminishes bias when comparing animal models.

Although this publication suggests that SVMs can be used to compare the distress of two animal models, it is premature to claim that this method will allow us to determine the severity of all animal models in a scientific and rational manner. First, distinct research facilities will have to test if this or similar methods can be applied to many different animal models to compare distress between distinct models. Second, accessible tools to assess and compare distress have to be provided for the scientific community. Talbot and colleagues have started to explore such a tool, and recommend the use of a Relative Severity Assessment (RELSA) score for comparing animal models[52]. It will be important for the research community to make such tools accessible online. Third, the scientific community will have to provide a network of comparing distress between the most essential animal models. Only if this network allows an arrangement of animal models according to their distress level, one could start grading evidence-based severity into categories (e.g. mild, moderate or severe) as demanded by the legislation of many countries.

Methods

Animals. This study was conducted in accordance with the European directive 2010/63/EU and national law. All experiments were approved by the local ethics committee of the public authority (Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei Mecklenburg-Vorpommern, 7221.3-1-002/17). Because female mice were used to expand the mouse strain, surplus male BALB/cANCrl mice were used for this study. Please note that the focus on male mice might be a limitation of this study. A few mice of this mouse strain were purchased from Charles River (Wilmington, MA USA) and bred in the central animal facility of the Rostock

University Medical Center (the health of the animal stock is routinely checked according to FELASA guidelines). Before the experiment the mice had more than 2 days for acclimatization. Animals were allocated in a non-random manner matching the age of both treatment groups and the experimenters were not blinded when injecting drugs. Distress was evaluated by two people (GT, NS), and in case of difficulties, in addition by another person (DZ). The required number of animals was calculated before starting the experiments by sample size calculation (alpha = 0.05, power = 0.8). Mice were group housed during breeding and the first few days before the actual experiments. Afterwards they were single housed in Eurostandard Type III clear plastic cages with wire lid, light/dark cycle of 12 h/12 h (dawn: 6:30–7:00 am) at a temperature of 21 ± 2 °C, with a relative humidity of 60 ± 20%. Autoclaved bedding (Bedding Espe Max 3–5 mm granulate, H 0234-500, Abedd, Vienna, Austria), shredded tissue paper (PZN03058052, FSMED Verbandmittel GmbH, Frankenberg, Deutschland), one paper tunnel (75 × 38 mm, H 0528-151, ssniff) and a wooden enrichment tool (Espe size S, 40 × 16 × 10 mm), H0234. NSG, Abedd). Food (pellets, V1534.000, 10 mm, ssniff) and tap water ad libitum were provided. Mice were euthanized by quickly anaesthetizing them with 5 vol % isoflurane and killing them with cervical dislocation.

Induction of liver damage. For inducing cholestasis by BDL on day 0, mice were quickly anaesthetized by 5 vol % isoflurane (CP-pharma, Burgdorf, Germany) and placed on a heating plate (37 °C). Then the laparotomy was performed under anesthesia (1.2–2.5 vol % isoflurane). As described in a previous study[53], the common bile duct was ligated by three surgical knots and was then transected between the two distal ligations. After closing the abdominal cavity, each mouse was allowed to recover from anesthesia in a single cage in front of a red warming lamp. The surgical procedure took 25–40 min. To relieve pain, 5 mg/kg carprofen (Pfizer GmbH, Berlin, Germany) was injected (sc) before operation and 0.25 ml metamizol (500 mg/ml, Ratiopharm GmbH, Ulm, Germany) was added to the drinking water (100 ml, drinking water was changed daily) until euthanasia of the mice. Supportive care was given after BDL by offering soaked food to all animals until euthanasia. In order to evaluate, if the NLRP3 inflammasome inhibitor MCC950 (Sigma Aldrich, St. Louis, USA, code PZ0280) could impair distress, 20 mg/kg MCC950 or aqua dest. ad inj. (Sham) was ip injected daily from day 1 before BDL to day 13 after BDL. For inducing liver damage by CCl_4 (Merck Millipore, Eschborn, Germany, code 1.02209.1000), this substance was diluted fourfold with corn oil (Sigma-Aldrich, code C8267). Per g body weight 1 µl of this solution (dose of CCl_4: 0.25 ml/kg body weight) was injected (ip) between 14:40–15:00 into the mice twice per week until day 42 (on day 0, 4, 7, 11, 14, 18, 21, 25, 28, 32, 35, 39). To relieve pain, 0.25 ml metamizol (500 mg/mL, Ratiopharm GmbH, Ulm, Germany) was added to the drinking water (100 ml) until euthanasia of the mice. 20 mg/kg MCC950 or aqua dest. ad inj. (Sham) was injected (ip) daily from day 28 to day 41 after first CCl_4 injection. The sixteen BDL mice (survivors) were at the beginning of the experiment 10.29/8.07–18.61 (median/interquartile range) weeks old and had 27.11/21.80–29.68 (median/interquartile range) g body weight, whereas ten BDL mice (non-survivors) were at the beginning of the experiment 9.79/8.36–12.20 weeks old and had 24.90/23.83–26.23 g body weight. The ten CCl_4-treated mice were at the beginning of the experiment 7.86/7.86–8.14 weeks old and had 24.52/22.99–24.97 g body weight.

Evaluation of animal distress. *Burrowing.* To evaluate burrowing activity of mice, a tube (length: 15 cm, diameter: 6.5 cm) filled with 200 g of food pellets was placed into the cage 2–3 h before the dark phase[54]. The remaining pellets in the burrowing tube were weighed after 17 ± 2 h and the weight of the burrowed pellets was calculated. Burrowing activity was measured before the first intervention (pre) and during the acute (day 0), early (BDL: day 1, CCl_4: day 4), middle (BDL: day 4, CCl_4: day 18) and late (BDL: day 13, CCl_4: day 39) phase of liver damage. The burrowing tube was always placed into the cage 1 ± 0.5 h after CCl_4 injection. Changes in burrowing activity were calculated by using the weight of burrowed pellets on day 7 before BDL and on day 8 before CCl_4 injection as a reference for the respective cohort.

Distress score. The wellbeing of mice was assessed by evaluating multiple parameters with the help of a distress score[55]. When the total score was higher than 15, the affected mouse was euthanized in order to avoid further deterioration of health. Distress was assessed before the first intervention (pre) and during the acute (day 0), early (BDL: day 1, CCl_4: day 4), middle (BDL: day 4, CCl_4: day 18) and late (BDL: day 13, CCl_4: day 39) phase of liver damage. The distress was always evaluated 30 ± 5 min after CCl_4 injection.

Body weight. The body weight of mice was assessed before the first intervention (pre) and during the acute (day 1), early (BDL: day 2, CCl_4: day 5), middle (BDL: day 5, CCl_4: day 19) and late (BDL: day 14, CCl_4: day 40) phase of liver damage. Thus, in all experiments the body weight was determined 1 day after measuring distress by a score sheet or by burrowing activity. This allows enough time for a body weight adjustment to a specific distress level (e.g. after injection of CCl_4).

Blood plasma and tissue analysis. AST, ALT, GLDH and ALP activity were spectrophotometrically assessed in blood plasma using the Cobas c111 analyser (Roche GmbH, Mannheim, Germany). For determining the corticosterone concentration in blood plasma the mouse and rat ELISA-Kit (DEV9922, Demeditec Diagnostics GmbH, Erfurt, Germany) was used according to the instructions of the manufacturer. Oxidative stress was evaluated by measuring the total malondialdehyde concentration after hydrolysing liver tissue at pH 1–2 and using the BIOXYTEC MDA-586 kit from OxisResearch (OXIS Health Products Inc. Portland, OR, USA).

Data presentation and statistical analysis. In line graphs data are presented as mean value ± standard deviation, whereas box plots indicate median interquartile range as well as 90% percentile and 10% percentile in form of

whiskers. The characteristics of data were assessed by Shapiro–Wilk normality test and by Levene median equal variance test. Student's t-test (based on normal distribution and equal variance of data) or the Mann–Whitney Rank Sum test were used to determine the significance of differences. When comparing two groups, differences with $P \leq 0.05$ were considered to be significant. When comparing treatment groups at several time points, differences were only considered to be significant, when the P-value was lower than 0.05 divided by the number of meaningful comparisons (Bonferroni correction for multiple comparisons). These evaluations were done using SigmaPlot 12.0 (SYSTAT Software Inc., San Jose, USA; https://systatsoftware.com/products/sigmaplot/). For box plots, ROC curves, logistic regressions and Support Vector Machine classification, data of the pre- and post-intervention phase (all data from the acute, early and middle phase) were used to differentiate between healthy and diseased animals. For differentiating between post-BDL survivors and non-survivors, all data of surviving mice of the acute, early and middle phase after BDL were compared to data measured 0–2 days before death or euthanasia of non-survivors.

ROC curve analysis (using SigmaPlot 12.0, SYSTAT Software Inc.) determined the area under the curve (AUC) with the respecting 95% confidence intervals (CI) as a measurement for the performance of the readout parameters[56]. In addition, this analysis gives the asymptotic P-value that determines if the AUC is significantly different from AUC = 0.5. To analyse the efficacy of the combination of two or three parameters, the data sets were combined by binary logistic regression using SigmaPlot 12.0 and the ROC curves were calculated afterwards.

In order to analyse distress considering all three readout parameters simultaneously, a Support Vector Machine was built on a 64-bit computer with 32 GB RAM using the R software[57] with the following packages: caret[58] and e1071[59]. Prior to model building, samples were class-labelled using the experimental time phases (pre- vs. post-intervention). Categories were labelled as level 0 (pre) and 1 (post) and used in the classification process. Samples were randomized into 70% training and 30% test data prior to model building. A linear kernel function (u'·v) was then used to construct the SVM-classifier with the training data. Data were scaled for the building process. The non-optimised fit was then tuned for the hyper-parameter cost function to optimise the SVM margin width for the classifier. In parallel, the tuning process was stratified using fivefold repeated tenfold cross-validation. The mean from all internal validation runs was then used to construct the optimised classifier. Model performance was reported in two stages: (a) re-classification (prediction) of the training data against the model (non-generalizable internal performance check) and (b) classification of the external test data (validation). In each case, data from a confusion matrix (accuracy, sensitivity, specificity) was reported for both, the optimised and the non-optimised model. The resulting values reflect model stability and also compensate for low sample sizes via repeated cross-validation. The externalised test data further assess the generalisability of the model. Finally, the hyperplane was constructed by coefficient extraction and grid extension of the optimised SVM model. When comparing CCl_4 cohorts to BDL, the optimized model was used to predict severity classes for post-intervention BDL data from the externalized test set as well as post-intervention CCl_4 data. The predictions were plotted in a scatterplot and class differences analyzed by Fisher's Exact Test.

Data availability

The authors declare that all data supporting the findings of this study are available within the paper and its supplementary information file.

Received: 12 November 2019; Accepted: 1 October 2020
Published online: 13 November 2020

References

1. 1*National Research Council (US) Committee for the Update of the Guide for the Care and Use of Laboratory Animals. Guide for the Care and Use of Laboratory Animals (Eighth Edition)*. (National Academy of Sciences, 2011).
2. 2Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes (Text with EEA relevance). Available from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:en:PDF. (2019).
3. Diaz, S. L. Conducting and reporting animal experimentation: Quo vadis?. *Eur. J. Neurosci.* https://doi.org/10.1111/ejn.14091 (2018).
4. Smith, A. J., Clutton, R. E., Lilley, E., Hansen, K. E. A. & Brattelid, T. PREPARE: Guidelines for planning animal research and testing. *Lab. Anim.* **52**, 135–141. https://doi.org/10.1177/0023677217724823 (2018).
5. 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**, e1000412. https://doi.org/10.1371/journal.pbio.1000412 (2010).
6. Smith, D. *et al.* Classification and reporting of severity experienced by animals used in scientific procedures: FELASA/ECLAM/ESLAV Working Group report. *Lab. Anim.* **52**, 5–57. https://doi.org/10.1177/0023677217744587 (2018).
7. 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**, 431–436 (1985).
8. Roughan, J. V. & Flecknell, P. A. Evaluation of a short duration behaviour-based post-operative pain scoring system in rats. *Eur. J. Pain* **7**, 397–406. https://doi.org/10.1016/S1090-3801(02)00140-4 (2003).
9. Graf, R., Cinelli, P. & Arras, M. Morbidity scoring after abdominal surgery. *Lab. Anim.* **50**, 453–458. https://doi.org/10.1177/0023677216675188 (2016).
10. Deacon, R. M. Burrowing in rodents: A sensitive method for detecting behavioral dysfunction. *Nat. Protoc.* **1**, 118–121. https://doi.org/10.1038/nprot.2006.19 (2006).
11. Jirkof, P. *et al.* Burrowing behavior as an indicator of post-laparotomy pain in mice. *Front. Behav. Neurosci.* **4**, 165. https://doi.org/10.3389/fnbeh.2010.00165 (2010).
12. Shepherd, A. J., Cloud, M. E., Cao, Y. Q. & Mohapatra, D. P. Deficits in burrowing behaviors are associated with mouse models of neuropathic but not inflammatory pain or migraine. *Front. Behav. Neurosci.* **12**, 124. https://doi.org/10.3389/fnbeh.2018.00124 (2018).
13. Lofgren, J. *et al.* Analgesics promote welfare and sustain tumour growth in orthotopic 4T1 and B16 mouse cancer models. *Lab. Anim.* **52**, 351–364. https://doi.org/10.1177/0023677217739934 (2018).

14. Hohlbaum, K. *et al.* Severity classification of repeated isoflurane anesthesia in C57BL/6JRj mice-assessing the degree of distress. *PLoS ONE* **12**, e0179588. https://doi.org/10.1371/journal.pone.0179588 (2017).
15. Hager, C. *et al.* Running in the wheel: Defining individual severity levels in mice. *PLoS Biol.* **16**, e2006159. https://doi.org/10.1371/journal.pbio.2006159 (2018).
16. Hajian-Tilaki, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian J. Intern. Med.* **4**, 627–635 (2013).
17. 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**, 1–13. https://doi.org/10.1258/la.2010.010031 (2011).
18. Baumans, V. Science-based assessment of animal welfare: Laboratory animals. *Rev. Sci. Tech.* **24**, 503–513 (2005).
19. Harikrishnan, V. S., Hansen, A. K., Abelson, K. S. & Sorensen, D. B. A comparison of various methods of blood sampling in mice and rats: Effects on animal welfare. *Lab. Anim.* **52**, 253–264. https://doi.org/10.1177/0023677217741332 (2018).
20. Peng, M. *et al.* Battery of behavioral tests in mice to study postoperative delirium. *Sci. Rep.* **6**, 29874. https://doi.org/10.1038/srep29874 (2016).
21. Moore, E. S. *et al.* Comparing phlebotomy by tail tip amputation, facial vein puncture, and tail vein incision in C57BL/6 mice by using physiologic and behavioral metrics of pain and distress. *J. Am. Assoc. Lab. Anim. Sci.* **56**, 307–317 (2017).
22. Hurst, J. L. & West, R. S. Taming anxiety in laboratory mice. *Nat. Methods* **7**, 825–826. https://doi.org/10.1038/nmeth.1500 (2010).
23. Kim, J. *et al.* Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19–9 blood markers. *Sci. Transl. Med.* https://doi.org/10.1126/scitranslmed.aah5583 (2017).
24. Booken, N. *et al.* Sezary syndrome is a unique cutaneous T-cell lymphoma as identified by an expanded gene signature including diagnostic marker molecules CDO1 and DNM3. *Leukemia* **22**, 393–399. https://doi.org/10.1038/sj.leu.2405044 (2008).
25. Guthrie, E. *et al.* Cluster analysis of symptoms and health seeking behaviour differentiates subgroups of patients with severe irritable bowel syndrome. *Gut* **52**, 1616–1622. https://doi.org/10.1136/gut.52.11.1616 (2003).
26. Giebeler, A. *et al.* c-Met confers protection against chronic liver tissue damage and fibrosis progression after bile duct ligation in mice. *Gastroenterology* **137**, 297–308. https://doi.org/10.1053/j.gastro.2009.01.068 (2009).
27. Modica, S. *et al.* Selective activation of nuclear bile acid receptor FXR in the intestine protects mice against cholestasis. *Gastroenterology* **142**, 355–356. https://doi.org/10.1053/j.gastro.2011.10.028 (2012).
28. Ding, B. S. *et al.* Divergent angiocrine signals from vascular niche balance liver regeneration and fibrosis. *Nature* **505**, 97–102. https://doi.org/10.1038/nature12681 (2014).
29. Scholten, D., Trebicka, J., Liedtke, C. & Weiskirchen, R. The carbon tetrachloride model in mice. *Lab. Anim.* **49**, 4–11. https://doi.org/10.1177/0023677215571192 (2015).
30. Cubero, F. J. *et al.* Combined activities of JNK1 and JNK2 in hepatocytes protect against toxic liver injury. *Gastroenterology* **150**, 968–981. https://doi.org/10.1053/j.gastro.2015.12.019 (2016).
31. Khan, N., Kuo, A., Brockman, D. A., Cooper, M. A. & Smith, M. T. Pharmacological inhibition of the NLRP3 inflammasome as a potential target for multiple sclerosis induced central neuropathic pain. *Inflammopharmacology* **26**, 77–86. https://doi.org/10.1007/s10787-017-0401-9 (2018).
32. Ghallab, A. *et al.* Influence of liver fibrosis on lobular zonation. *Cells* **8**, 1556. https://doi.org/10.3390/cells8121556 (2019).
33. Bleich, A. & Tolba, R. H. How can we assess their suffering? German research consortium aims at defining a severity assessment framework for laboratory animals. *Lab. Anim.* **51**, 667. https://doi.org/10.1177/0023677217733010 (2017).
34. Ezure, T. *et al.* The development and compensation of biliary cirrhosis in interleukin-6-deficient mice. *Am. J. Pathol.* **156**, 1627–1639. https://doi.org/10.1016/S0002-9440(10)65034-1 (2000).
35. Georgiev, P. *et al.* Characterization of time-related changes after experimental bile duct ligation. *Br. J. Surg.* **95**, 646–656. https://doi.org/10.1002/bjs.6050 (2008).
36. Gabele, E. *et al.* TNFalpha is required for cholestasis-induced liver fibrosis in the mouse. *Biochem. Biophys. Res. Commun.* **378**, 348–353. https://doi.org/10.1016/j.bbrc.2008.10.155 (2009).
37. Yi, H. S. *et al.* Treatment with 4-methylpyrazole modulated stellate cells and natural killer cells and ameliorated liver fibrosis in mice. *PLoS ONE* **10**, e0127946. https://doi.org/10.1371/journal.pone.0127946 (2015).
38. Yoshioka, H. *et al.* Vitamin D3-induced hypercalcemia increases carbon tetrachloride-induced hepatotoxicity through elevated oxidative stress in mice. *PLoS ONE* **12**, e0176524. https://doi.org/10.1371/journal.pone.0176524 (2017).
39. Munafo, M. R. & Davey Smith, G. Robust research needs many lines of evidence. *Nature* **553**, 399–401. https://doi.org/10.1038/d41586-018-01023-3 (2018).
40. Habib, A. *et al.* Inhibition of monoacylglycerol lipase, an anti-inflammatory and antifibrogenic strategy in the liver. *Gut* https://doi.org/10.1136/gutjnl-2018-316137 (2018).
41. Zhang, K. *et al.* The liver-enriched lnc-LFAR1 promotes liver fibrosis by activating TGFbeta and Notch pathways. *Nat. Commun.* **8**, 144. https://doi.org/10.1038/s41467-017-00204-4 (2017).
42. Yanguas, S. C. *et al.* Experimental models of liver fibrosis. *Arch. Toxicol.* **90**, 1025–1048. https://doi.org/10.1007/s00204-015-1543-4 (2016).
43. Liedtke, C. *et al.* Experimental liver fibrosis research: Update on animal models, legal issues and translational aspects. *Fibrogen. Tissue Repair* **6**, 19. https://doi.org/10.1186/1755-1536-6-19 (2013).
44. Manibusan, M. K., Odin, M. & Eastmond, D. A. Postulated carbon tetrachloride mode of action: A review. *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* **25**, 185–209. https://doi.org/10.1080/10590500701569398 (2007).
45. Secklehner, J. & Richardson, C. A. The reporting of animal welfare details in liver research: A review of studies describing bile duct ligation in mice (2011–2013). *J. Hepatol.* **62**, 250–251. https://doi.org/10.1016/j.jhep.2014.09.029 (2015).
46. Liles, J. H. & Flecknell, P. A. The influence of buprenorphine or bupivacaine on the post-operative effects of laparotomy and bile-duct ligation in rats. *Lab. Anim.* **27**, 374–380. https://doi.org/10.1258/002367793780745552 (1993).
47. Jirkof, P. Side effects of pain and analgesia in animal experimentation. *Lab. Anim. (NY)* **46**, 123–128. https://doi.org/10.1038/laban.1216 (2017).
48. Jirkof, P. *et al.* Administration of tramadol or buprenorphine via the drinking water for post-operative analgesia in a mouse-osteotomy model. *Sci. Rep.* **9**, 10749. https://doi.org/10.1038/s41598-019-47186-5 (2019).
49. Ovadia, C. *et al.* Association of adverse perinatal outcomes of intrahepatic cholestasis of pregnancy with biochemical markers: Results of aggregate and individual patient data meta-analyses. *Lancet* **393**, 899–909. https://doi.org/10.1016/S0140-6736(18)31877-4 (2019).
50. Molinaro, A. M. Diagnostic tests: How to estimate the positive predictive value. *Neurooncol. Pract.* **2**, 162–166. https://doi.org/10.1093/nop/npv030 (2015).
51. Moller, C. *et al.* Toward evidence-based severity assessment in rat models with repeated seizures: I. Electrical kindling. *Epilepsia* **59**, 765–777. https://doi.org/10.1111/epi.14028 (2018).
52. Talbot, S. R. *et al.* One score to rule them all: severity assessment in laboratory mice. *bioRxiv* https://doi.org/10.1101/2020.06.23.166801 (2020).
53. Abshagen, K. *et al.* Pathobiochemical signatures of cholestatic liver disease in bile duct ligated mice. *BMC Syst. Biol.* **9**, 83. https://doi.org/10.1186/s12918-015-0229-0 (2015).
54. Deacon, R. Assessing burrowing, nest construction, and hoarding in mice. *J. Vis. Exp.* https://doi.org/10.3791/2607 (2012).

55. Kumstel, S. *et al.* Grading distress of different animal models for gastrointestinal diseases based on plasma corticosterone kinetics. *Animals (Basel)* https://doi.org/10.3390/ani9040145 (2019).
56. Bewick, V., Cheek, L. & Ball, J. Statistics review 13: Receiver operating characteristic curves. *Crit. Care* **8**, 508–512. https://doi.org/10.1186/cc3000 (2004).
57. R Core Team. The R project for statistical computing. https://www.R-project.org/. (2019).
58. Kuhn, M. Caret: Classification and regression training. R package version 6.0–84. https://CRAN.R-project.org/package=caret. (2019).
59. David, M., Evgenia, D., Kurt, H., Andreas, W. & Leisch., F. e1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. R package version 1.7–2. https://CRAN.R-project.org/package=e1071. (2019).

Acknowledgements

Guanglin Tang and Xianbin Zhang were supported by the China Scholarship Council (Grant number: 201808080167 and 201608080159). The study was supported by the Deutsche Forschungsgemeinschaft (DFG research group for 2591, Grant number: 321137804, ZE 712/1-1, and VO 450/15-1).

Author contributions

Study design: D.Z., B.V. and K.A.; Data collection and analysis or interpretation: all authors; Drafting manuscript: G.T., S.R.T., X.Z., and D.Z.; Revising manuscript: all authors. Approved final version of manuscript: all authors. G.T, S.R.T., X.Z. and D.Z. are responsible for the integrity of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-76391-w.

Correspondence and requests for materials should be addressed to S.R.T., X.Z. or D.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ALTEX preprint
published December 9, 2019
doi:10.14573/altex.1909111

Research Article

A rational approach of early humane endpoint determination in a murine model for cholestasis

Xianbin Zhang[1], Simone Kumstel[1*], Guanglin Tang[1*], Steven R. Talbot[2], Nico Seume[1], Kerstin Abshagen[1], Brigitte Vollmar[1] and Dietmar Zechner[1]*

[1]Rudolf-Zenker-Institute of Experimental Surgery, University Medical Center Rostock, Rostock, Germany; [2]Institute for Laboratory Animal Science, Hannover Medical School, Hannover, Germany

Abstract

Reduction of animal suffering during *in vivo* experiments is usually ensured by continuously monitoring the health status using a score sheet and by applying humane endpoints. However, most studies do not evaluate the plausibility of score sheets and do not attempt to reduce the suffering of animals by determining earlier and therefore more humane endpoints. The present study uses data assessed from BALB/cANCrl mice after bile duct ligation to retrospectively analyze which score sheet criteria are informative to determine humane endpoints. The performance of each single as well as combinations of multiple animal welfare parameters was analyzed by a Cox proportional-hazards model followed by Harrell's concordance index. The addition of behavioral parameters, such as burrowing activity, helped to define a more humane early endpoint for euthanizing these animals. Using this approach, we determined that a body weight loss of 10-20 % combined with a reduction of burrowing activity by more than 79.4 % was able to predict that these animals died within two days. Thus, this approach successfully determined an earlier humane endpoint and will therefore reduce the suffering of animals in future experiments. A consequent application of such an approach or similar methods will contribute to the refinement of various animal experiments.

1 Introduction

According to animal protection laws enacted by most nations (EU, 2010; Germany, 2013) high animal welfare standards are a prerequisite to get the permission to perform animal based research. Moreover, these standards also provide an important foundation for high quality of biomedical research (Bayne and Würbel, 2014; Carbone and Austin, 2016). Thus, it is in the interest of the public community and scientists to alleviate suffering of animals.

One key aspect to reduce animal suffering is to determine humane endpoints for timely euthanasia. Criteria for humane endpoints are, for example, 20 % body weight loss (Morton, 2000) and hypothermia or lethargy (Acred et al., 1994). However, these symptoms often reflect severe suffering just before death. Defining more "humane" criteria, which are able to predict death within a certain period of time, could reduce the suffering of laboratory animals. As an ideal tool to grade suffering of animals and to determine humane endpoints, a clinical score sheet or "welfare assessment protocol" was established by Morton and Griffiths in 1985 (Morton and Griffiths, 1985). The score sheets should include reasonable criteria to recognize pain, suffering or discomfort of the animals during the specific procedures. The common criteria are body weight loss, appearance, spontaneous and flight behavior as well as intervention specific clinical signs. According to these criteria the distress of an animal can be ranged in mild, moderate and severe (Morton and Griffiths, 1985; Hawkins et al., 2011; Smith et al., 2018).

Many different scoring systems were already established for different animal models (Paster et al., 2009; Kanzler et al., 2016) and animal welfare organizations published many helpful protocols to constantly improve the current score sheets (Hawkins et al., 2011; Smith et al., 2018). These organizations recommend also the use of behavioral parameters to analyze the psychological state of the animals in addition to physical criteria such as body weight, posture, body temperature etc. (Hawkins et al., 2011). Thus, well-being in rodents is also often assessed by analyzing natural behaviors such as burrowing and nesting activity (Deacon, 2006a,b; Jirkof et al., 2013b; Jirkof, 2014). This proved to be useful to detect neurological, abdominal and post-surgical pain or stress in mice and rats (Deacon et al., 2005; Jirkof et al., 2013a,b; Jirkof, 2014; Pfeiffenberger et al., 2015; Sliepen et al., 2019; Jirkof et al., 2010).

*contributed equally

Received September 11, 2019; Accepted November 27, 2019;
Epub December 9, 2019; © The Authors, 2019.

ALTEX 37(#), ###-###. doi:10.14573/altex.1909111

Correspondence: PD Dr. Dietmar Zechner,
Rudolf-Zenker-Institute for Experimental Surgery, University Medical Center Rostock,
Schillingallee 69a, 18057 Rostock, Germany
(dietmar.zechner@uni-rostock.de)

Although many different scoring systems have been published world-wide to determine animal suffering and to provide a basis for euthanasia of animals, very few publications have attempted to define which read-out parameters are informative (Nemzek et al., 2004; Mai et al., 2018; Leung et al., 2019) and even fewer publications describe methods to optimize their scoring system in order to determine early humane endpoints (Nunamaker et al., 2013; Koch et al., 2016).

Thus, this study assessed multiple animal welfare parameters on mice, which underwent bile duct ligation (BDL), a widely used procedure to study liver damage and fibrosis. We critically evaluate a scoring system in order to optimize it by measuring animal behavior and further, to determine more humane endpoint criteria.

2 Material and Methods

2.1 Animals

For this study we used a total of 55 male, 10/9.6-13.1 weeks (median/interquartile range) old BALB/cANCrl mice, with an average body weight of 25.6/ 23.9-27.4 g (median/ interquartile range). Breeding pairs were originally purchased from Charles River and further bred in the facility of the University Medical Center in Rostock under specific pathogen free conditions. The mice had an acclimatization time of more than 2 days before the experiments started. During the experiment the mice were kept single housed in Eurostandard Type III plastic cages (Zoonlab GmbH, Castrop-Rauxel, Germany) with a light dark cycle of 12h/12h at a temperature of 21 ± 2 °C (dawn: 6:30-7:00 am) and a relative humidity of 60 ± 20 %. Food (pellets, 10 mm, ssniff-Spezialdiäten GmbH, Soest, Germany) and tap water were provided *ad libitum*. Enrichment was supplied by nesting material as shredded tissue paper (PZN03058052, FSMED Verbandmittel GmbH, Frankenberg, Germany), one paper tunnel (75 × 38 mm, H 0528-151, ssniff) and a wooden enrichment tool (Espe size S, 40 × 16 × 10 mm, H0234.NSG, Abedd, Vienna, Austria). Due to low sociability and high aggression of male BALB/cANCrl mice (Brodkin, 2007; Jones and Brain, 1987), the animals were single housed during the experiments. The animal experiment was approved by the local ethics committee and public authority (Landesamt für Landwirtschaft, Lebensmittelsicherheit und Fischerei Mecklenburg-Vorpommern, 7221.3-1-002/17), in accordance with the European directive 2010/63/EU as well as the national law of Germany and is reported according to the ARRIVE Guidelines (Kilkenny et al., 2010).

2.2 Induction of liver damage

For the induction of cholestasis by BDL the mice were anesthetized by 1.2-2.5 vol. % isoflurane (CP-pharma, Burgdorf, Germany), and placed on a heating plate at 37 °C in the laboratory. Isoflurane was chosen because it allowed a fast recovery from anesthesia. 5 mg/kg carprofen (Rimadyl®, Pfizer GmbH, Berlin, Germany) was injected subcutaneously 5-15 minutes before the start of the surgical intervention for perioperative analgesia. The eyes were kept wet by eye ointment. A midline laparotomy was performed and the bile duct was ligated three times with 5-0 silk and transected between the two distal ligations (Abshagen et al., 2015). The peritoneum and the skin were closed separately by 6-0 and 4-0 prolene suture (Johnson & Johnson MEDICAL GmbH, New Brunswick) and the mice were placed in front of a heating lamp. The surgical procedure lasted 25-40 min. Wet pellets (10 mm, ssniff-Spezialdiäten GmbH) were provided as refinement during the first days of recovery. 1250 mg/L Metamizol (Ratiopharm, Ulm, Germany) was administered via drinking water through the whole experiment for pain relief/management. In order to evaluate a possible therapeutic efficacy of the NLRP3 inflammasome inhibitor MCC950 (Sigma Aldrich, St. Louise, USA), 20 mg/kg MCC950 or aqua (control) was injected daily between 8:00-10:00 am intraperitoneally from day -1 before BDL to day 13 after BDL. Animals were allocated in a non-random manner matching the age of both treatment groups and the researchers were not blinded when injecting drugs. The mice were euthanized by cervical dislocation after a short anesthesia by 5 vol. % isoflurane on day 14 after BDL or when one of the humane endpoint criteria was met according to the attached score sheet (Tab. 1).

2.3 Cohorts of mice and assessment of distress

From 55 mice, 20 were euthanized or did not survive until day 14 after BDL. These mice were defined as non-survivors and were at the beginning of the experiment 10.0/8.9-12.1 weeks (median/interquartile range) old. These mice had a body weight of 25.0/23.8-26.6 g (median/interquartile range). In addition, 35 mice survived until day 14 after BDL and were therefore defined as survivors, with an age of 11.9/10.1-13.7 (median/interquartile range) weeks and a weight of 26.7/24.3-27.8 g (median/interquartile range). We observed that 75 % (15/20) of non-survivors were euthanized or died within 4 days after BDL. Thus, these mice died within 3.0/2.0-4.75 (median/interquartile range) days. Therefore, we measured distress on day 1 and day 4 after BDL for the survivor and non-survivor cohort. In addition, for those non-survivors from which we could not obtain the data on day 4, we adopted the data points measured 0-2 days before death or euthanasia.

Distress of 55 mice (35 survivors = 70 data points; 20 non-survivors = 33 data points since 7 mice had to be euthanized, before the second score could be assessed; this results in 103 data points) was evaluated by a score sheet (Tab. 1 and Tab. S1[1]). This score-sheet was based on previously published score sheets (Morton, D. and Griffiths, 1985; Paster et al., 2009) and was already used in our working group to evaluate murine animal models for gastrointestinal diseases (Kumstel et al., 2019). The distress score was assessed between 8:30-10:30 am in the home cage. According to the defined score sheet, body weight, appearance, spontaneous and flight behavior as well as intervention specific clinical signs were assessed in a non-blinded fashion by two people (GT, NS) and in case of discrepancies by a third observer (DZ).

Burrowing behavior (variable VI) of 24 mice (16 survivors = 32 data points; 8 non-survivors= 13 data points, 3 mice had to be euthanized before the second data point was assessed, this results in 45 data points) was analyzed according to Deacon et al. (Deacon, 2006b). A tube was filled with 200 g pellets (10 mm, ssniff-Spezialdiäten GmbH) and placed into the home cage 2.5-3 h before the dark phase (16:00-16:30) and the burrowed amount of pellets was calculated 17 h later.

[1] doi:10.14573/altex.1909111s

Tab. 1: Score-Sheet
(Kumstel et al., 2019)

Observations (variables)	Score
I Body weight	
I-a decreased > 10% (compared to initial weight)	2
I-b decreased > 20% (compared to initial weight)	5
II General condition	
II-a tooth displacement, too long teeth	1 (A)
II-b fur dull, ruffled or untended	2
II-c eyes unclear or squinted	2
II-d untended orifices of the body	3
II-e abnormal posture	3
II-f dehydration	3
II-g short spasms or temporary paralysis symptoms	3
II-h persistent (>30') cramping or paralysis	5
II-i abnormal respiratory sounds or animal feels cold	5
III Spontaneous behavior	
III-a the animal is passive or overactive	2
III-b pronounced apathy, hyperkinetic, or isolation	4
III-c squeaking due to pain	5
III-d self-mutilation	5
IV Flight behavior after contact	
IV-a animal is passive or overactive	2
IV-b distinct apathy or hyperkinetic	5
V Process-specific criteria	
V-a wound healing disorder	2
V-b opening of the sutures by biting	1 (B)
V-c local inflammation	2
V-d ascites	4
Total score	0-66

Nesting activity (variable VII) was assessed on a different cohort of animals than burrowing activity, because both assessments might influence each other. Nesting activity was evaluated on 22 mice (15 survivors = 30 data points; 7 non-survivors = 10 data points since 4 mice did not survive long enough, this results in 40 data points). Nesting activity was assessed by providing a nestlet in the home cage (5 cm square of pressed cotton batting, Zoonlab GmbH, Castrop-Rauxel, Germany) 0.5-1 h before dark phase (18:00-18:30). The nest was scored at the next morning (9:00-11:00 am) according to a 1-5 point scale of Deacon. (Deacon, 2006a). We additionally scored 6 points for a perfect nest, when more than 90 % of the circumference of the walls was higher than the mouse. To enable individual learning, both behavior tests were performed three times in group housing until the mice were housed separately.

Body weight reduction was assessed on 48 mice (35 survivors = 70 data points; 13 non-survivors = 21 data points since 5 mice died before the second data point for bodyweight reduction could be assessed; this results in 91 data points).

2.4 Development of an optimal prognosis model

Continuous variables, such as reduction of burrowing and nesting activity, were converted into dichotomous variables according to Youden's index (Ruopp et al., 2008). Based on this approach the best cut-off value to distinguish between survivors and non-survivors was at 79.4 % reduction of burrowing behavior or a nesting score of less than 2.5. The Kaplan-Meier estimator followed by log-rank test was performed by SigmaPlot 12.0 (SYSTAT Software Inc., San Jose, USA), and all variables, which could significantly discriminate between the survival time of non-survivors and survivors, were used to develop the prognosis model by univariate and multivariate Cox proportional-hazards model (SigmaPlot 12.0, SYSTAT Software Inc., San Jose, USA). In addition, to evaluate whether the proportional hazards assumption is satisfied, we performed Log-minus-log plots, a frequently used method for the validation of a proportional hazard assumption (In and Lee, 2018), for all variables used for strategy 1, strategy 2 and strategy 3. To determine the optimal prognosis model, the Harrell's concordance indices (C-indices) were investigated using the R software (Foundation for Statistical Computing, Vienna, Austria) with the help of the Hmisc (Harrell, 2019) and survival packages (Liu, P. et al., 2017; Therneau, 2019). C-indices were internally validated by 10000-fold bootstrapping of the parametrized survival models using the boot package (Canty and Ripley, 2019; Davison and Hinkley, 1997). Resulting means and the bias-corrected and accelerated (BCa) 95 % confidence intervals (CI) were reported.

3 Results

3.1 Retrospective analysis of distress parameters

After removing distress parameters (II-a, II-d, II-g, II-h, III-c, III-d, V-a, V-b, V-c), which were never observed during the experiment and parameters (I-b, II-i, IV-b), which demanded immediate euthanasia eleven variables were evaluated by Kaplan-Meier estimator (Fig. 1). There was no significant difference regarding the survival time when considering some single variables, such as criteria II-b (fur dull, ruffled or untended), II-e (abnormal posture) and V-d (ascites) (Fig. 2). However, mice with a positive status of I-a (body weight loss of 10 % to 20 %), II-c (eyes unclear or squinted), II-f (dehydration), III-a (spontaneous behavior: passive or overactive), III-b (pronounced apathy, hyperkinetic, or isolation), IV-a (passive or overactive

ALTEX preprint
published December 9, 2019
doi:10.14573/altex.1909111

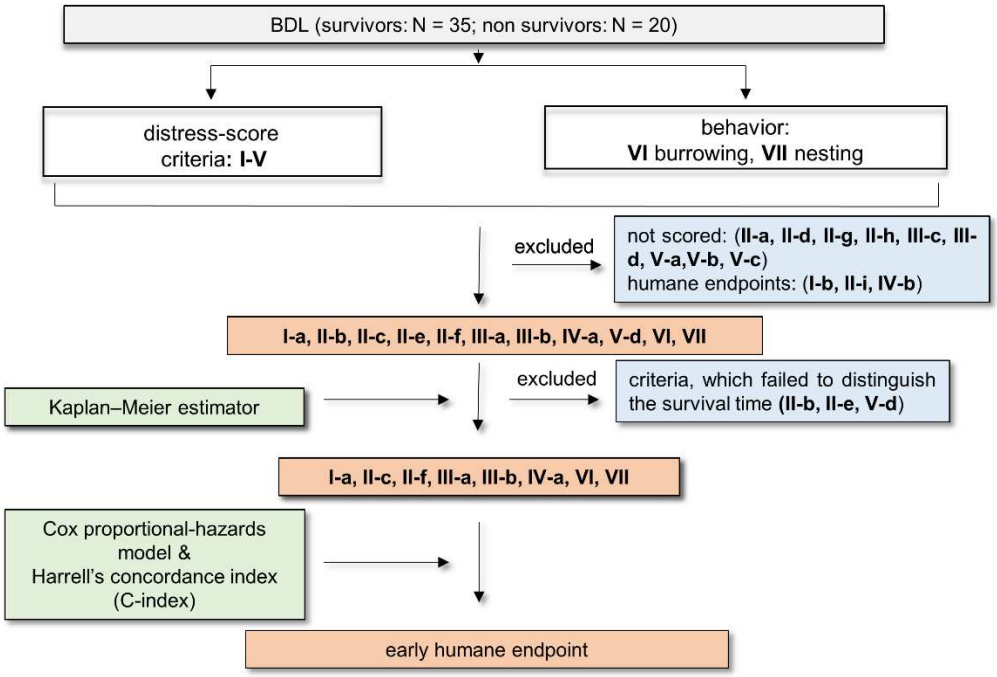


Fig. 1: Flowchart to retrospectively analyze distress parameters in order to determine an early humane endpoint
First, score sheet criteria (for details see Tab. 1), which were not observed during the experiment as well as humane endpoint criteria, were excluded. Second, Kaplan-Meier estimator curves were used to exclude criteria, which did not predict survival time. Third, the performance of each single as well as combinations of multiple parameters were analyzed by Cox proportional-hazards model followed by Harrell's concordance index, to determine which criteria combination might be used as efficient early humane endpoint.

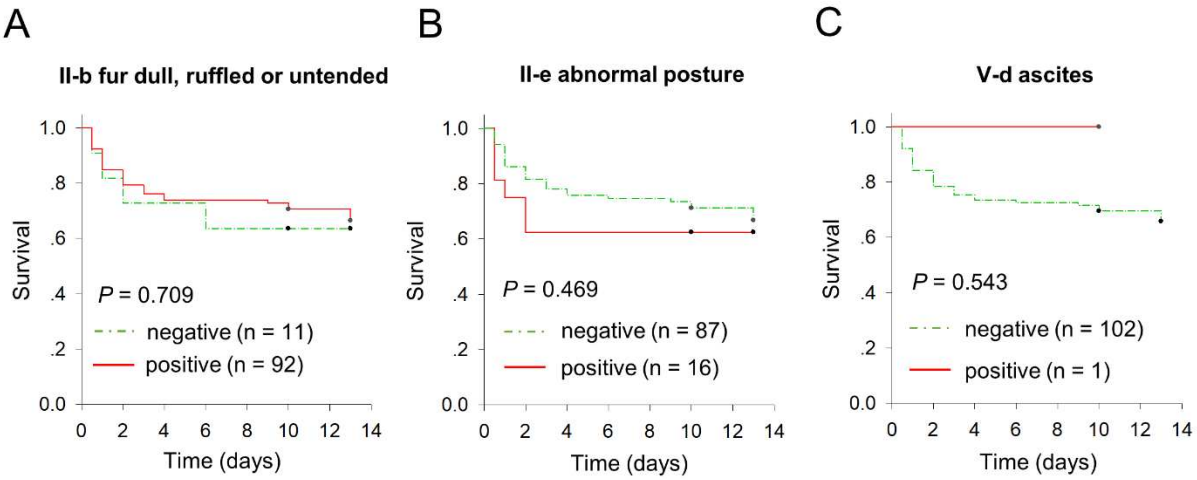


Fig. 2: Kaplan-Meier curves using distinct score sheet criteria
The positive status of II-b (fur dull, ruffled or untended) (**A**), II-e (abnormal posture) (**B**) or V-d (ascites) (**C**) failed to significantly indicate reduced survival time of mice. The *P*-value was determined by log-rank test.

after being touched by the observer), VI (reduction of burrowing activity more than 79.4 %), or VII (decrease of nesting score more than 2.5) had a significantly ($P \leq 0.005$) shorter survival time than mice with negative status of these criteria (Fig. 3). Interestingly, all mice with a positive status of I-a had to be euthanized within 14 days and mice with pronounced apathy (III-b) or which were passive or overactive after being touched by the researcher (IV-a) had to be euthanized within 1-2 days. This suggests that I-a, III-b and IV-a might be powerful criteria to define early humane endpoints for the BDL model.

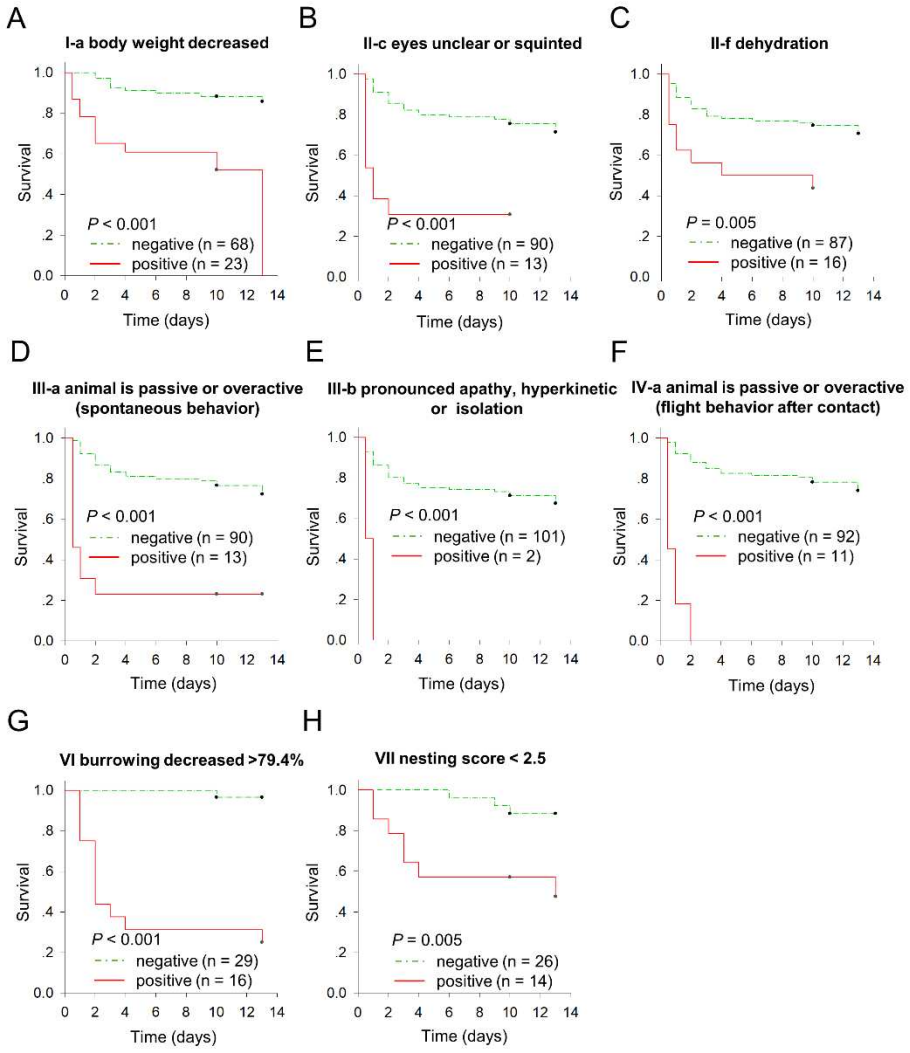


Fig. 3: Kaplan-Meier curves using distinct read-out parameters for distress
For the positive status of I-a (body weight decreased 10 to 20 %) (**A**), II-c (eyes unclear and squinted) (**B**), II-f (dehydration) (**C**), III-a (spontaneous behavior: animal is passive and over active) (**D**), III-b (pronounced apathy, hyperkinetic or isolation) (**E**), IV-a (flight behavior after contact: animal is passive or overactive;) (**F**), decreased burrowing activity by > 79.4 % (**G**) and nesting score of less than 2.5 (**H**), a significantly shorter survival time was calculated by log rank test ($P \leq 0.005$).

3.2 Predictive models for defining survival time

We first used a univariate Cox proportional-hazards model in order to evaluate the hazard ratio (HR) of each variable. We observed that a positive status of I-a, II-c, II-f, III-a, III-b, IV-a, VI or VII significantly increased the risk of death when compared to the negative status of these variables (Tab. 2). We then performed a multivariate analysis of these distress score criteria (strategy 1). Two variables, I-a and IV-a were found to significantly affect the hazard rate. Thus the positive status of these two variables significantly increased the risk of death (strategy 1 in Tab. 2). We then included two behavioral parameters in our analysis, burrowing and nesting activity. Since these two variables were assessed in two distinct cohorts, we developed two additional multivariate Cox proportional-hazards models (strategy 2 and strategy 3). First, we added burrowing activity to all distress score variables (strategy 2). We observed that the positive status of variable I-a (body weight) and VI (burrowing) significantly increased the risk of death (strategy 2 in Tab. 2). We also added nesting activity to all distress score variables (strategy 3). We observed that a multivariate Cox proportional-hazards model only recognized that nesting activity significantly increased the risk of death (strategy 3 in Tab. 2). The following observations support the concept that the conditions for applying the COX proportional hazard model are met. First, we do no observe that curves in the Kaplan-Meier curves cross. In addition, we performed Log-minus-log plots for all variables used for strategy 1 (Fig. S1[1]), strategy 2 (Fig. S2[1]) and strategy 3 (Fig. S3[1], for strategies see Tab. 2) and observed that these curves are parallel.

3.3 Evaluation of models for defining survival time

Models, which use single variables or combinations of variables were compared to each other using Harrell's Concordance index. We also bootstrapped the indices of each model for getting more robust estimates, thereby internally validating the goodness-of-fit (Tab. 3) of each model. We observed that model 2 (C-index: 0.947, 95 % CI: 0.832-0.978) has a higher C-

Tab. 2: HR (hazard ratio) and P-value (P) for the distinct variables were determined by Cox proportional-hazards model

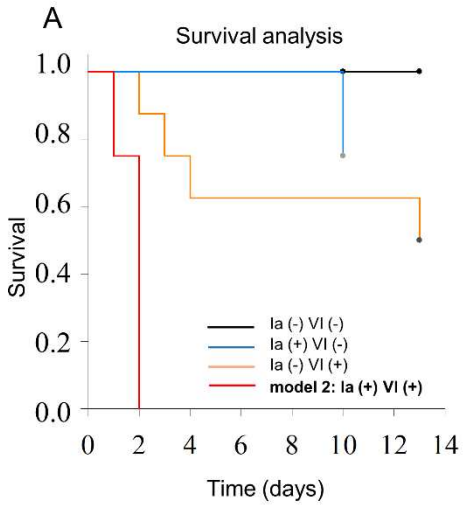
Variables		Univariate		Multivariate					
		HR	*P*	Strategy 1		Strategy 2		Strategy 3	
				HR	*P*	HR	*P*	HR	*P*
I-a	negative	reference		reference		reference		NS	
	positive	6.169	< 0.001	3.968	0.017	24.096	0.007		
II-c	negative	reference		NS		NS		-	
	positive	5.033	< 0.001						
II-f	negative	reference		NS		NS		NS	
	positive	2.825	0.009						
III-a	negative	reference		NS		NS		NS	
	positive	6.584	< 0.001						
III-b	negative	reference		NS		-		-	
	positive	9.495	0.003						
IV-a	negative	reference		reference		NS		-	
	positive	17.130	< 0.001	8.621	0.021				
VI	negative	reference		-		reference		-	
	positive	33.218	< 0.001			54.348	0.003		
VII	negative	reference		-		-		reference	
	positive	5.639	0.013					5.051	0.026

NS indicates no significant difference (P > 0.05).

Tab. 3: Bootstrapped C-indices (Harrell's concordance index) and the corresponding 95 % confidence intervals for each distinct model

	Variables	C-index	C-index bootstrapped	95% CI
Single variables	I-a	0.691	0.692	0.572-0.797
	II-c	0.633	0.632	0.553-0.721
	II-f	0.590	0.590	0.517-0.680
	III-a	0.657	0.656	0.574-0.746
	III-b	0.534	0.534	0.500-0.614
	IV-a	0.684	0.683	0.603-0.770
	VI	0.865	0.866	0.768-0.923
	VII	0.725	0.724	0.534-0.848
Model-1	Ia, IVa	0.719	0.720	0.590-0.838
Model-2	**Ia, VI**	**0.943**	**0.947**	**0.832-0.978**
Model-3	VII	0.725	0.724	0.532-0.849
Model-4	I-a, II-c, II-f, III-a, III-b, IV-a	0.696	0.726	0.553-0.782

index than model 1 (C-index: 0.720, 95 % CI: 0.590-0.838) or model 3 (C-index: 0.724, 95 % CI: 0.532-0.849). In addition, we also evaluated model 4 (C-index: 0.726, 95 % CI: 0.553-0.782), where we used a combination of all distress score parameters, which significantly increased the risk to die based on univariate Cox proportional-hazards model. We observed that model 2 also had a higher C-index than model 4 (Tab. 3). The C-index of model 2, body weight plus burrowing activity (C-index: 0.947, 95 % CI: 0.832-0.978), was also higher than that of each single variable (body weight C-index: 0.692, 95 % CI: 0.572-0.797; burrowing activity weight C-index: 0.866, 95 % CI: 0.768-0.923), as well as the C-index of all other single variables (Tab. 2). Please note that a C-index of 0.5 corresponds to a non-informative prediction rule, whereas a C-index of 1 corresponds to a perfect prediction rule (Schmid et al., 2016). Since the C-index as well as the bootstrapped C-index of our model 2 is above 0.9, this suggests that model 2 can very well predict the survival times of these animals. In order to evaluate the practicability of using this combination, we evaluated it by Kaplan-Meier estimator. We observed that mice which lost body weight in a range of 10 % to 20 % and also showed decreased burrowing activity by more than 79.4 % had a significantly shorter survival time when compared to mice which were positive for only one or none of these two variables (Fig. 4). All mice with such a reduction in body weight and burrowing activity died within 2 days (Fig. 4). In addition, we analyzed, if age had an influence on the humane endpoint. We determined the optimal cut-off of age (10.21 weeks) by Youden's index and observed that young mice (age < 10.21 weeks) had a significant reduction of survival time when compared to elderly mice (Fig. S4A[1]). We compared how our best model (using variable I-a plus variable VI) predicts death in young or old mice and observed that this model could predict death of old mice within one day (Fig. S4B[1]) and death of young mice within two days (Fig. S4C[1]).



B

P-values	Ia (-) VI(-)	Ia (+) VI(-)	Ia (-) VI(+)
Ia (-) VI(-)	-	-	-
Ia (+) VI(-)	0.025	-	NS
Ia (-) VI(+)	0.002	NS	-
Ia (+) VI(+)	< 0.001	0.030	0.014

Fig. 4: Kaplan–Meier curve of the multivariate Model 2
Mice with a body weight loss of more than 10 to 20 % (variable I-a) and decreased burrowing activity of more than 79.4 % (variable VI) died within 2 days (**A**). Mice positive for variable Ia and VI had a significantly decreased survival time when compared to mice positive for just one variable or negative for both variables (**B**). The *P*-values were determined by log-rank test ($P < 0.05$).

4 Discussion

In the current study, we present an approach of how to retrospectively analyze a score sheet and how to include additional distress parameters in order to define an early humane endpoint. Mice which lost 10-20 % body weight and reduced their burrowing activity by more than 79.4 % died within 2 days. Thus, applying these cut-off criteria for an earlier euthanasia might reduce the suffering of animals in subsequent experiments. A consequent application of such an approach in most animal experiments might contribute to the refinement of animal research.

This approach is based on three steps (Fig. 1). First, score sheet criteria, which were not observed during the experiment were excluded. Second, Kaplan-Meier estimator curves helped to exclude criteria, which did not contribute in predicting survival times. Third, the performance of each single as well as combinations of multiple parameters were further analyzed by a Cox proportional-hazards model followed by Harrell's concordance index.

Interestingly, survival analysis using the Kaplan-Meier estimator suggested that the single variables III-b (pronounced apathy, hyperkinetic or isolation) and IV-a (animal is passive or overactive- flight behavior after contact) are able to predict death within 1-2 days (Fig. 3E and F). Therefore, one could consider these two criteria for determining an early humane endpoint. However, the incidence of variable III-b (n=1) was rather low. Moreover, the accuracy according to the C-indices (III-b: 0.534; IV-a: 0.684) of these two criteria were also quite low when compared to the multivariate model 2 (C-index: 0.943). This indicates that a multivariate distress analysis can be superior to univariate analysis. This conclusion is consistent with other studies, which also suggest that using more than one parameter for death prediction is beneficial (Hankenson et al., 2013; Trammell and Toth, 2011; Ray et al., 2010).

Our multivariate approach was implemented by Cox proportional-hazard model and Harrell's concordance index of models. These methods are commonly applied in statistical medical research to investigate the survival time of patient's reliant on one or more independent variables. Using this approach, we were able to define distinct univariate and multivariate models. A comparison of the C-indices revealed that the combination of body weight loss and reduction of burrowing behavior is the best model for predicting survival times (Tab. 3). An internal validation of the goodness-of-fit via bootstrapping the C-indices of each model also demonstrated the robustness of model 2. Model 2 has a C-index well above 0.9. A value of C = 0.5

corresponds to a non-informative prediction rule, whereas C = 1 corresponds to a perfect prediction rule (Schmid et al., 2016). In many meaningful biomedical applications, Harrel's Concordance index often ranges between the values 0.6 and 0.75. This was reported, for example, by Van Belle et al. (2011), Schröder et al., (2011) and Zhang et al. (2013). Since both the C-index and the bootstrapped C-index of our model 2 is above 0.9, we conclude that model 2 can very well predict the survival times of these animals. The two variables body weight and burrowing behavior have the additional advantage that they can be objectively measured, which minimizes potential selection bias. It is also well-accepted that both criteria reliably measure suffering of mice (Jirkof et al., 2013b; Deacon et al., 2005; Pfeiffenberger et al., 2015; Hohlbaum et al., 2017; Häger et al., 2018). However, it is likely that for each animal model a different combination of parameters might be best to predict survival times and to determine early humane endpoints.

This study relied on the evaluation of physical parameters, i.e. score sheet criteria and body weight or behavioral parameters such as burrowing as well as nesting activity for determining a humane endpoint. Body weight has been demonstrated to be very useful for determining humane endpoints by many other studies (Takayama-Ito et al., 2017; Trammell and Toth, 2011; Hankenson et al., 2013). The advantage of body weight is that it is applicable as humane endpoint to many species and it is easy to assess. However, for cancer studies this parameter proved to be less useful as exclusive indicator for euthanasia (Paster et al., 2009). Since the adaption of body weight after an intervention or during a disease takes about 24 h, this endpoint criterion is also not practical for acute and severe diseases. In contrast to body weight loss, only one unsuccessful attempt was published so far, where burrowing activity was used to determine a humane endpoint in a chemotherapy induced mucositis model in rats (Whittaker et al., 2015). However, burrowing behavior is known to be a sensitive indicator for distress after surgical interventions (Jirkof et al., 2010; Jirkof et al., 2013a) or during chronic diseases (Jirkof, 2013b; Abdelrahman et al., 2019). Even the suffering of neurological disorders (Deacon et al., 2005; Felton et al., 2005), or very mild stress inductions such as isoflurane aesthesia (Hohlbaum et al., 2017) can be quantified by burrowing behavior. A disadvantage of burrowing activity is that it is only applicable for rodents. Subjective score sheet criteria, i.e. cramping and paralysis, abnormal respiratory sounds, squeaking due to pain, self-mutilation or apathy and hyperkinetic were also introduced in other studies as humane endpoints for mice and rats (Kanzler et al., 2016; Brabb et al., 2014; Herrmann and Flecknell, 2018). Concerning bile duct ligation in mice the following criteria were defined so far as humane endpoint: distension of abdomen, ascites, debilitating diarrhea, bleeding from the orifices, peritonitis, internal bleeding or sepsis (Tag et al., 2015a,b). These humane endpoints mostly describe pathologies instead of symptoms. The presented score sheet criteria focus on symptoms which might be caused by these pathologies.

Some studies used body temperature as an important parameter for determining humane endpoints (Mai et al., 2018; Mei et al., 2018; Warn et al., 2003; Drechsler et al., 2015; Kort et al., 1998; Cates et al., 2014). Body temperature is considered to be the most accurate humane endpoint criterion, especially for acute and severe infections (Nemzek et al., 2004; Adamson et al., 2013). Thus, the lack of an accurate measurement of body temperature might be a limitation of this study. A major drop in body temperature is observed in sepsis models and acute infections a few hours before death (Napier et al., 2016; Mai et al., 2018; Kort et al., 1998). Since it was our goal to develop early humane endpoints, we abstained from measuring the body temperature. Another limitation of this study might be that we determined an earlier humane endpoint by using only data of day 1 and day 4 after BDL. To address this point we added additional data taken at day 2 (from 8 non-survivors and 35 survivors) and analyzed survival by Kaplan Meyer curves. This resulted in very similar survival curves as shown in Fig. 3, suggesting that additional data have little influence on our results (data not shown). Since we could demonstrate that elderly mice (>10.21 weeks) have a better survival than young mice, using elderly mice will reduce the number of needed animals and might help to minimize the distress caused by this animal model.

Some studies have also used multivariate analysis for early humane endpoint determination. (Trammell and Toth, 2011; Nunamaker et al., 2013). However, these studies did not compare the accuracy of single criteria or multiple combinations for predicting death. By using the combination of up to three parameters, i.e. body weight, temperature and a neuroscore an elaborate machine learning approach for early humane endpoint determination in mouse models for stroke and sepsis was established (Mei et al., 2019). The accuracy of this model (stroke: 0.93; sepsis: 0.96) proved to be quite high. However, the machine learning tool needs to be trained with physical data from the specific animal model and a large sample size is necessary to build a generalized model. Hankenson et al. established an early humane endpoint model using the combination of body temperature (< 34.5 °C) and weight loss (more than 0.05 g daily) by linear regression (Hankenson et al., 2013). Linear regression estimates diagnostic outcomes at the moment of prediction, while Cox regression determines prognostic outcomes within a distinct period of time (Moons et al., 2015). The inclusion of time as a factor is an advantage of our approach, because it also provides the information of how fast an animal dies. Thus, compared to the already published methods for early humane endpoint determination the advantage of our approach is the inclusion of the time variable by Cox regression and that a justifiable low number of animals is required for analysis.

References

Abdelrahman, A., Kumstel, S., Zhang, X., Liebig, M. et al. (2019). A novel multi-parametric analysis of non-invasive methods to assess animal distress during chronic pancreatitis. *Sci Rep 9*, 14084. doi:10.1038/s41598-019-50682-3

Abshagen, K., König, M., Hoppe, A., Müller, I. et al. (2015). Pathobiochemical signatures of cholestatic liver disease in bile duct ligated mice. *BMC Syst Biol 9*, 83. doi:10.1186/s12918-015-0229-0

Acred, P., Hennessey, T. D., MacArthur-Clark, J. A., Merrikin, D. J. et al. (1994). Guidelines for the welfare of animals in rodent protection tests. A report from the Rodent Protection Test Working Party. *Lab Anim 28*, 13–18. doi:10.1258/002367794781065870

Adamson, T. W., Diaz-Arevalo, D., Gonzalez, T. M., Liu, X. et al. (2013). Hypothermic Endpoint for an Intranasal Invasive Pulmonary Aspergillosis Mouse Model. *Comp Med 63* (6), 477–481.

Bayne, K., Würbel, H. (2014). The impact of environmental enrichment on the outcome variability and scientific validity of laboratory animal studies. *Rev - Off Int Epizoot 33* (1), 273–280. doi:10.20506/rst.33.1.2282

Brabb, T., Carbone, L., Snyder, J., Phillips, N. (2014). Institutional animal care and use committee considerations for animal models of peripheral neuropathy. *ILAR J 54*, 329–337. doi:10.1093/ilar/ilt045

Brodkin, E. S. (2007). BALB/c mice: low sociability and other phenotypes that may be relevant to autism. *Behav Brain Res 176*, 53–65. doi:10.1016/j.bbr.2006.06.025

Canty, A., Ripley, B. (2019). boot: Bootstrap R (S-Plus) Funtions. R package version. Version 1.3-22.

Carbone, L., Austin, J. (2016). Pain and Laboratory Animals: Publication Practices for Better Data Reproducibility and Better Animal Welfare. *PLoS ONE 11*, e0155001. doi:10.1371/journal.pone.0155001

Cates, C. C., McCabe, J. G., Lawson, G. W., Couto, M. A. (2014). Core Body Temperature as Adjunct to Endpoint Determination in Murine Median Lethal Dose Testing of Rattlesnake Venom. *Comp Med 64* (6), 440–447.

Davison, A.C.; Hinkley, D.V. (1997). Bootstrap methods and their application. Cambridge: Cambridge University Press (Cambridge series on statistical and probabilistic mathematics, 1). doi:10.1017/CBO9780511802843

Deacon, R. M. J. (2006a). Assessing nest building in mice. *Nat Protoc 1*, 1117–1119. doi:10.1038/nprot.2006.170

Deacon, R. M. J. (2006b). Burrowing in rodents: a sensitive method for detecting behavioral dysfunction. *Nat Protoc 1*, 118–121. doi:10.1038/nprot.2006.19

Deacon, R. M. J., Reisel, D., Perry, V. H., Nicholas, J. et al. (2005). Hippocampal scrapie infection impairs operant DRL performance in mice. *Behav Brain Res 157*, 99–105. doi:10.1016/j.bbr.2004.06.013

Drechsler, S., Weixelbaumer, K. M., Weidinger, A., Raeven, P. et al. (2015). Why do they die? Comparison of selected aspects of organ injury and dysfunction in mice surviving and dying in acute abdominal sepsis. *ICMx 3*, 1247. doi:10.1186/s40635-015-0048-z

EU - 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 of September 2010 on the protection of animals used for scientific purposes. 2010/63/EU. *Offical Journal of the European Union* (L-276), 33–79. doi:10.1007/978-1-137-54482-7_33

Felton, L. M., Cunningham, C., Rankine, E. L., Waters, S. et al. (2005). MCP-1 and murine prion disease: separation of early behavioural dysfunction from overt clinical disease. *Neurobiol Dis 20*, 283–295. doi:10.1016/j.nbd.2005.03.008

Germany (2013). Tierschutzgesetz, TierSchG, vom 04.07.2013 (BGBl.S1950). In: BGBl. http://www.gesetze-im-internet.de/tierschg/BJNR012770972.html, accessed on November, 2019.

Häger, C., Keubler, L. M., Talbot, S. R., Biernot, S. et al. (2018). Running in the wheel: Defining individual severity levels in mice. *PLoS Biol 16*, e2006159. doi:10.1371/journal.pbio.2006159

Hankenson, F. C., Ruskoski, N., van Saun, M., Ying, G.-S. et al. (2013). Weight Loss and Reduced Body Temperature Determine Humane Endpoints in a Mouse Model of Ocular Herpesvirus Infection. *J Am Assoc Lab Anim Sci 52* (3), 277–285.

Harrell, F. (2019). Package 'Hmisc'. https://cran.r-project.org/web/packages/survival/survival.pdf, accessed on August, 2019.

Hawkins, P., Morton, D. B., Burman, O., Dennison, N. 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. doi:10.1258/la.2010.010031

Herrmann, K., Flecknell, P. (2018). Severity classification of surgical procedures and application of health monitoring strategies in animal research proposals: A retrospective review. *Altern Lab Anim 46* (5), 273–289. doi:10.1177/026119291804600606

Hohlbaum, K., Bert, B., Dietze, S., Palme, R. et al. (2017). Severity classification of repeated isoflurane anesthesia in C57BL/6JRj mice-Assessing the degree of distress. *PLoS ONE 12*, e0179588. doi:10.1371/journal.pone.0179588

In, J., Lee, D. K. (2018). Survival analysis: Part I - analysis of time-to-event. *Korean J Anesthesiol 71*, 182–191. doi:10.4097/kja.d.18.00067

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

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

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

Jirkof, P., Leucht, K., Cesarovic, N., Caj, M. et al. (2013). Burrowing is a sensitive behavioural assay for monitoring general wellbeing during dextran sulfate sodium colitis in laboratory mice. *Lab Anim 47*, 274–283. doi:10.1177/0023677213493409

Jones, S. E., Brain, P. F. (1987). Performances of inbred and outbred laboratory mice in putative tests of aggression. *Behav Genet 17*, 87–96. doi:10.1007/BF01066013

Kanzler, S., Rix, A., Czigany, Z., Tanaka, H. et al. (2016). Recommendation for severity assessment following liver resection and liver transplantation in rats: Part I. *Lab Anim 50*, 459–467. doi:10.1177/0023677216678018

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

Koch, A., Gulani, J., King, G., Hieber, K. et al. (2016). Establishment of Early Endpoints in Mouse Total-Body Irradiation Model. *PLoS ONE 11*, e0161079. doi:10.1371/journal.pone.0161079

Kort, W. J., Hekking-Weijma, J. M., TenKate, M. T., Sorm, V. et al. (1998). A microchip implant system as a method to determine body temperature of terminally ill rats and mice. *Lab Anim 32*, 260–269. doi:10.1258/002367798780559329

Kumstel, S., Tang, G., Zhang, X., Kerndl, H. et al. (2019). Grading Distress of Different Animal Models for Gastrointestinal Diseases Based on Plasma Corticosterone Kinetics. *Animals 9*, 145. doi:10.3390/ani9040145

Leung, V. S.Y., Benoit-Biancamano, M.-O., Pang, D. S.J. (2019). Performance of behavioral assays: the Rat Grimace Scale, burrowing activity and a composite behavior score to identify visceral pain in an acute and chronic colitis model. *Pain Rep 4*. doi:10.1097/PR9.0000000000000712

Liu, P., Zhang, X., Shang, Y., Lu, L. et al. (2017). Lymph node ratio, but not the total number of examined lymph nodes or lymph node metastasis, is a predictor of overall survival for pancreatic neuroendocrine neoplasms after surgical resection. *Oncotarget 8*, 89245–89255. doi:10.18632/oncotarget.19184

Mai, S. H. C., Sharma, N., Kwong, A. C., Dwivedi, D. J. et al. (2018). Body temperature and mouse scoring systems as surrogate markers of death in cecal ligation and puncture sepsis. *Intensive Care Med Exp 6*, 20. doi:10.1186/s40635-018-0184-3

Mei, J., Banneke, S., Lips, J., Kuffner, M. T. C. et al. (2019). Refining humane endpoints in mouse models of disease by systematic review and machine learning-based endpoint definition. *ALTEX*. doi:10.14573/altex.1812231

Mei, J., Riedel, N., Grittner, U., Endres, M. et al. (2018). Body temperature measurement in mice during acute illness: implantable temperature transponder versus surface infrared thermometry. *Sci Rep 8*, 3526. doi:10.1038/s41598-018-22020-6

Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A. et al. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med 162*, W1-73. doi:10.7326/M14-0698

Morton, D., Griffiths, P. (1985). Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *Veterinary Record 116*, 431–436. doi:10.1136/vr.116.16.431

Morton, D. B. (2000). A systematic approach for establishing humane endpoints. *ILAR J 41*, 80–86. doi:10.1093/ilar.41.2.80

Napier, B. A., Brubaker, S. W., Sweeney, T. E., Monette, P. et al. (2016). Complement pathway amplifies caspase-11–dependent cell death and endotoxin-induced sepsis severity. *J. Exp. Med. 213*, 2365–2382. doi:10.1084/jem.20160027

Nemzek, J. A., Xiao, H.-Y., Minard, A. E., Bolgos, G. L. et al. (2004). Humane endpoints in shock research. *Shock 21*, 17–25. doi:10.1097/01.shk.0000101667.49265.fd

Nunamaker, E. A., Artwohl, J. E., Anderson, R. J., Fortman, J. D. (2013). Endpoint Refinement for Total Body Irradiation of C57BL/6 Mice. *Comp Med 63* (1), 22–28.

Paster, E. V., Villines, K. A., Hickman, D. L. (2009). Endpoints for Mouse Abdominal Tumor Models: Refinement of Current Criteria. *Comp Med 59* (3), 234–241.

Pfeiffenberger, U., Yau, T., Fink, D., Tichy, A. et al. (2015). Assessment and refinement of intra-bone marrow transplantation in mice. *Lab Anim 49*, 121–131. doi:10.1177/0023677214559627

Ray, M. A., Johnston, N. A., Verhulst, S., Trammell, R. A. et al. (2010). Identification of Markers for Imminent Death in Mice used in Longevity and Aging Research. *J Am Assoc Lab Anim Sci 49* (3), 282–288.

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., Schisterman, E. F. (2008). Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biom J 50*, 419–430. doi:10.1002/bimj.200710415

Schmid, M., Wright, M., Ziegler, A. (2016). On the use of Harrell's C for clinical risk prediction via random survival forest. http://arxiv.org/pdf/1507.03092.pdf, accessed on November, 2019.

Schröder, M. S., Culhane, A. C., Quackenbush, J., Haibe-Kains, B. (2011). survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics 27*, 3206–3208. doi:10.1093/bioinformatics/btr511

Sliepen, S. H. J., Diaz-Delcastillo, M., Korioth, J., Olsen, R. B. et al. (2019). Cancer-induced Bone Pain Impairs Burrowing Behaviour in Mouse and Rat. *In Vivo 33*, 1125–1132. doi:10.21873/invivo.11582

Smith, D., Anderson, D., Degryse, A.-D., Bol, C. 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. doi:10.1177/0023677217744587

Tag, C. G., Sauer-Lehnen, S., Weiskirchen, S., Borkham-Kamphorst, E. et al. (2015a). Bile duct ligation in mice: induction of inflammatory liver injury and fibrosis by obstructive cholestasis. *J Vis Exp*. doi:10.3791/52438

Tag, C. G., Weiskirchen, S., Hittatiya, K., Tacke, F. et al. (2015b). Induction of experimental obstructive cholestasis in mice. *Lab Anim 49*, 70–80. doi:10.1177/0023677214567748

Takayama-Ito, M., Lim, C.-K., Nakamichi, K., Kakiuchi, S. et al. (2017). Reduction of animal suffering in rabies vaccine potency testing by introduction of humane endpoints. *Biologicals 46*, 38–45. doi:10.1016/j.biologicals.2016.12.007

Therneau, T. (2019). Package 'survial'. https://cran.r-project.org/web/packages/survival/survival.pdf, accessed on August, 2019.

Trammell, R. A., Toth, L. A. (2011). Markers for Predicting Death as an Outcome for Mice Used in Infectious Disease Research. *Comp Med 61* (6), 492–498.

van Belle, V., Pelckmans, K., van Huffel, S., Suykens, J. A. K. (2011). Improved performance on high-dimensional survival data by application of Survival-SVM. *Bioinformatics 27*, 87–94. doi:10.1093/bioinformatics/btq617

Warn, P. A., Brampton, M. W., Sharp, A., Morrissey, G. et al. (2003). Infrared body temperature measurement of mice as an early predictor of death in experimental fungal infections. *Lab Anim 37*, 126–131. doi:10.1258/00236770360563769

Whittaker, A. L., Lymn, K. A., Nicholson, A., Howarth, G. S. (2015). The assessment of general well-being using spontaneous burrowing behaviour in a short-term model of chemotherapy-induced mucositis in the rat. *Lab Anim 49*, 30–39. doi:10.1177/0023677214546913

Zhang, H., Xia, W., Lu, X., Sun, R. et al. (2013). A novel statistical prognostic score model that includes serum CXCL5 levels and clinical classification predicts risk of disease progression and survival of nasopharyngeal carcinoma patients. *PLoS ONE 8*, e57830. doi:10.1371/journal.pone.0057830

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft (DFG research group FOR 2591, project number: 321137804, ZE 712/1-1 and VO 450/15-1). Xianbin Zhang and Guanglin Tang were supported by the China Scholarship Council (grant numbers: 201608080159 and 201808080167).