| UTRECHT UNIVERSITY |

| FACULTY OF VETERINARY MEDICINE |

| DEPARTMENT OF ANIMALS IN SCIENCE AND SOCIETY |

MASTER THESIS

NUMBERS BECOME INDIVIDUALS

WHETHER AN IMPROVED CONTROL OVER INTER-INDIVIDUAL DIFFERENCES CAN CONTRIBUTE TO THE QUALITY OF ANIMAL EXPERIMENTS

(1)

S. Stolte BSc
E-mail: s.stolte@uu.nl
Student number: 5596378
Supervisor: Drs. M.H. van der Goot
In collaboration with M.H. Kooij BSc (student nr.: 5644518)
4-7-2019

ABSTRACT

A major predicament when doing behavioural research are the dissimilarities found between individuals, regardless of measures that have been taken to ensure standardisation within the experiment. In the light of the three R's (replacement, reduction and refinement) regarding laboratory animal testing, means to control and/or identify these inter-individual differences may improve the reliability and reproducibility of research and may lead to a reduction of the required sample size in the future. The present study focused on comparing a study design that attempted to control the inter-individual variation vs. a study design that did not take such variation into consideration.

59 BALB/cAnNCrl mice, 60 C57BL/6NCrl mice and 60 129S2/PasCrl mice were used in the experiment. All mice were subjected to behavioural testing in a modified Hole Board and blood samples were taken from each respective mouse to analyse corticosterone levels. Afterwards, cluster analysis was run on the behavioural and corticosterone data to allocate the mice into different response type clusters. In the next phase, dexmedetomidine was used as treatment and the emotional reactivity from the previous phase was used to compose experimental groups.

It was revealed that there were two cluster response types within the group of tested animals, with distributions of 52.5% and 47.5%. Moreover, different significant results and effect sizes in the group balanced in response type cluster in comparison to the group unbalanced in response type cluster.

Overall, the present study observed that taking inter-individual variation of intersession habituation into consideration when creating experimental groups was deemed to improve the reliability and repeatability of the study. The presented method offers flexibility and can thus be readily applied in other behavioural studies.

Keywords:
inter-individual variation, k-means clustering, modified Hole Board, emotional reactivity, dexmedetomidine.

TABLE OF CONTENTS

1. INTRODUCTION AND AIM

In modern-day preclinical research, genetic uniformity and environmental standardisation are used as means to increase the reliability of studies (2,3). In order for research to be reliable, it should have been validated through, for example, replication (3,4). Since research on anxiety-related disorders often demands the use of mice and rats as animal models (5–7), mouse and rat inbred models that facilitate genetic uniformity have been established (3), thus making validation by standardising achievable. Regardless of these methods of standardisation, however, substantial diversity in physiological and behavioural response between different individuals of the same strain has been demonstrated to exist (8–12). As a result, the inter-individual variation can potentially impact the reliability of research results (13) by decreasing chances of validation through replication (14). In the current time and age, increasing the sample size is a commonly used approach to be able to disregard the variation between individuals (11,13).

Determining the cause of the diversity in behavioural and physiological response between individual experimental animals has proven to be complicated, seeing that recent studies have shown a variety of different explanations for said diversity, ranging from (epi)genetic and environmental factors (9,15–17) to a dissimilarity in intestinal flora (18,19). The results of these studies are mostly consistent on one conclusion: the diversity in behavioural and physiological response between individuals of the same population is caused by complex gene by environment interaction (9,15,16,18,19), even in the case of genetic uniformity between said individuals (12). As a result, it is difficult to regulate inter-individual variation through standardisation.

Identifying and/or controlling the different types of responses, behavioural as well as physiological, may thus lead to an improvement in quality of animal experiments. However, few scientific articles have been published on the subject of using statistics to take this diversity into account rather than establishing its cause (20). Improving identification in diversity between animals of the same strain may decrease the sample size needed in experiments (14), promoting reduction of the required amount of laboratory animals and keeping the 3 R's (replacement, reduction and refinement) (21) in mind.

As part of ongoing research on inter-individual differences, the present study was designed to evaluate whether an improved control over inter-individual differences in mice can contribute to the quality of animal experiments. Prior to this study, retrospective analyses and an experiment have been conducted to demonstrate and validate the existence of inter-individual differences. This master thesis study focused on the comparison between a study design that attempted to control the inter-individual variation and a study design that did not take inter-individual variation into account.

1.1 RETROSPECTIVE ANALYSES AND PRIOR EXPERIMENT

Before the start of the present study, retrospective analyses attempting to explore the existence of inter-individual differences (22) have been performed on previously published studies (23,24). These studies focused on adaptive capacities in male BALB/cJ, 129P3/J, 129P2/OlaHsd, 129S2/SvHsd, 129S2/SvPasCrl, 129X1 and 129P2/J mouse inbred strains. Moreover, all studies used a modified Hole Board set-up to show difference in habituation/adaptation patterns between the different (sub)strains (23,24). To analyse the data from these studies and possibly identify inter-individual differences, pattern analysis had been applied using a combination of linear mixed models and clustering techniques (22,25). Pattern analysis was able to yield proof on the existence of inter-individual variation, but also opened up the possibility to map this variation. The retrospective pattern analysis resulted in three clusters as optimal cluster solutions, dubbed type A "Non-responding", type B "Sensitisation" and type C "Habituation".

The different analysed studies lacked consistency, i.e. the retrospective data was collected from different experiments that were done in different time frames, by different researchers, on different locations, et cetera (22). As inconsistencies existed between the analysed studies, the results and method (i.e. the k-means clustering) were in need of validation. An experiment (Exp. 1) has been conducted prior to this study to explore, validate and extend the aforementioned results in further detail. In this experiment, BALB/cAnNCrl, C57BL/6NCrl and 129S2/PasCrl mice were subjected to the modified Hole Board, during which behavioural analysis as well as examination of corticosterone levels was carried out. Exp. 1 validated the results from the retrospective data, again by applying the k-means clustering method and by yielding three clusters as optimal cluster solutions: response type A "Non-responding", response type B "Sensitisation" and response type C "Habituation". Additionally, this experiment demonstrated that it is possible to identify different response types, whereby the types vary in the degree and approach in which the animals adapted to the test set-up. Some strains seemed to lean more towards a specific response type, as was illustrated by BALB/c mice that dominated type C. On the other hand, the results from C57BL/6 and 129S2 mice seemed to be spread out more evenly across types A and B. (26)

The present study, defined as Exp. 2, will investigate whether taking inter-individual variation into consideration contributes to the standardisation of behavioural studies.

1.2 BACKGROUND OF THE MODIFIED HOLE BOARD

The modified Hole Board (mHB) has been utilised in Exp. 1 and the retrospective studies (26) and was used again in the present study (Exp. 2) to facilitate behavioural scoring of mice. The mHB combines characteristics of both the hole board and the open field test, resulting in a paradigm that allows for a larger variety of behavioural dimensions in rodents to be studied in one single test (27,28). It consists of a board that has multiple holes in its centre, which is placed in the centre of a box, thus creating an open field in the centre of said box where the holes are (27). The broad behavioural spectrum of rodents (28) can be evaluated with more ease due to the combination of the open field test's characteristic leeway for locomotor activity, defecation, risk-assessment and avoidance behaviour (29), the mHB's inherent possibility to insert a novel object recognition test (28), the hole board test's evocation of locomotion and (directed) exploration (29) and the mHB's circumvention of needing socially isolated and/or food deprived animals (28). In lieu of using a multiple test battery, as was done in the past (e.g. 34,35), the mHB decreases the required sample size and reduces the amount of time and money that is needed to effectively execute a multiple test battery (27).

Altogether, the mHB allows for a wide variety of behavioural parameters to be tested. Due to this large amount of parameters, it was more efficient to organise the variables in composite variables (32). Composite variables are variables that are the result of the combination of multiple correlating variables (32). There is a variety of different ways to accomplish combining different valuables, but the present study made use of z-scoring. Guilloux et al. (2011) described z-scoring regarding behavioural analysis in mice, which showed how "*integrating measures along the same behavioural dimensions in different tests would reduce the intrinsic variability of single tests and provide a robust characterization of the underlying "emotionality" of individual mouse*" (33). Through z-transformation, z-scores for specific variables can be created and combined to form a composite z-score (32). In the present study, it was decided to organise the behavioural variables into five different behavioural dimensions: avoidance behaviour, risk-assessment, locomotion, exploration and arousal. The five dimensions have previously been established through principal component analysis by Laarakker and colleagues (34) and have been confirmed with the use of z-scoring (35). All dimensions have been validated as reliable on multiple occasions by other studies (23,36–38) (for more information on the five dimensions, see: 3.4 Modified Hole Board and z-scoring).

1.3 MOUSE STRAINS AND THEIR BACKGROUND

The mouse strains that were used in the present study were BALB/cAnNCrl, C57BL/6NCrl and 129S2/PasCrl, meaning that the mice were all inbred. The advantage of using inbred strains of mice was that fewer animals were required to generate reliable and specific results (39). Moreover, using inbred strains means that the present study is more easily replicated (39).

As the BALB/c strain had also been used in the retrospective analyses and Exp. 1, it has been used again to collect data for this master thesis. The BALB/c strain consists of white mice. It is one of the most used mice strains, together with C57BL/6 mice (39). Consequently, the use of the BALB/c strain and the C57BL/6 strain in the present study is of even greater importance, due to the fact that mapping their inter-individual diversity can be relevant to a great variety of other studies. Furthermore, the BALB/c mice are commonly used in research into cancer and immunology and the strain shows genetic intactness, meaning that the progenitor strain has not been subjected to genetic mixing with any other strain in the past (40). The behaviour of BALB/c mice has also been described to be more “emotional” (41–43), meaning that the behavioural and physiological changes due to a conscious mental reaction (44) of these mice are much more prominent. Next, the 129S2 substrain was chosen to partake in this study, also due to the 129 strain being a part of the retrospectively analysed studies and Exp. 1. This agouti coated mouse strain is used most commonly in research to generate transgenic and knockout mouse lines (45) and has demonstrated a maladaptive habituation pattern (23,24). Lastly, the C57BL/6 strain had not been used in the data of the retrospective analyses, but was used in this study and Exp. 1 due to being the most widely used strain in biomedical research (46). Its popularity in behavioural studies in particular stems from the traits of the black mouse strain: the mice are capable of learning a variety of tasks, breed frequently and are physically active (46). In contrast to the BALB/c strain, the C57BL/6 mice have been described as “non-emotional” (41–43), meaning that they experience less of a behavioural and physiological change due to conscious mental reactions (44).

Regarding the aforementioned parameters in the mHB, preceding studies have revealed several inter-strain differences concerning the BALB/c and the C57BL/6 strains in an mHB setting. For example, C57BL/6 mice were revealed to be more active in terms of locomotor activity (38). This coincides with the results of other studies that were also on inter-strain differences in locomotor activity, although these studies did not make use of the mHB paradigm (37,47,48). Furthermore, the BALB/c mice were reported to be more active than C57BL/6 mice in terms of directed exploration in a hole board test (37), although a similar significant difference has yet to be found using the mHB (38).

2. RESEARCH QUESTION

A previous experiment (Exp. 1) attempted and succeeded to validate the results found in retrospective analyses on the existence of a pattern in different individual responses, which resulted in the discovery of three response types (22). Consequently, the present study (Exp. 2) investigated if identifying said response types in mice could contribute to the quality of animal experiments. The following is thus the hypothesis of this study: "*Taking inter-individual variation in emotional reactivity into account in composition of experimental groups of mice improves reliability of results*."

To properly examine the probability of this hypothesis, Exp. 1 (see: 1.1 Retrospective analyses and prior experiment) has first validated a way to identify the existence of inter-individual variation in emotional reactivity within three strains. The exact same three strains (BALB/cAnNCrl, C57BL/6NCrl and 129S2/PasCrl) have been used again in Exp. 2. In the first phase of the present study, *n* = 179 mice have been subjected to the mHB in five consecutive trials to examine their intersession habituative response and to establish response type clusters on this population. Moreover, during this phase, blood samples were taken seven days before testing with the mHB, on the day of testing (immediately after testing) and seven days after testing to measure corticosterone levels.

In phase 2, the different strains of mice were separately divided into subgroups: a treatment group and a control group. An anxiolytic (dexmedetomidine) was used on the treatment group, whereas a placebo (saline) was used on the control group. The groups were then split further into two groups: one group was balanced in terms of response types, while the other group was unbalanced in terms of response types. To clarify, the balanced groups consisted of animals with the same response type (which lead to six groups) and the unbalanced groups had mixes of response types in the animals (which lead to six groups as well). These balanced/unbalanced groups both had treatment and control groups (meaning three balanced and three unbalanced per treatment/placebo group) and, using the mHB once, the responses between these four groups were compared. As there were three strains that were subjected to testing, two different types of treatment (dexmedetomidine and saline) and two main groups (balanced/non-balanced with respect to response type), this lead to a total of 12 different groups (3x2x2 design).

Since taking response type cluster into account is a means to improve standardisation, it was established that if the results were affected by considering response type cluster, reliability and repeatability through standardisation would be deemed enhanced. Therefore, the hypothesis was confirmed if a difference between balanced and unbalanced group results within a behavioural dimension were found.

Altogether, this research examined the inter-strain differences, the effects of dexmedetomidine in mice, the effects of balanced experimental groups derived from response type on reliability of research and the interactions between these variables. As the main goal of this master thesis was to assess the effects of controlling inter-individual differences on reliability of research, the inter-strain differences and effects of dexmedetomidine were investigated as well, since they possibly contained valuable supplementary results.

3. MATERIALS AND METHODS

3.1 ETHICAL NOTE

The protocols of the experimental phases have been approved by the Central Animal Experiments Committee (CCD license number: AVD1080020172264). The resolution of this approval was reached on the basis of the EU directive 2010/63/EU (Directive on the Protection of Animals used for Scientific Purposes). The experiments have been conducted according to the 'Dutch Code on Laboratory Animal Car and Welfare' and have been reported according to the ARRIVE guidelines (49).

3.2 THE PRESENT EXPERIMENT AND ITS DIFFERENT PHASES

As a follow-up experiment to Exp. 1 (see: 1.1 Retrospective analyses and prior experiment), the present study examined if the use of individual variation in habituation profiles affects the variation in read-out parameters and thus the quality of animal research. To be able to test this, inbred males of the three different mouse strains were characterised on their response type cluster in phase 1. All animals were in the same experimental group and were tested using the mHB. Corticosterone levels were also taken into account.

After the characterisation in phase 1, a pilot was executed to determine the required dose of dexmedetomidine. Next, phase 2 tested if the use of response types resulted in a better quality of research by dividing the animals into experimental groups that were balanced by response type and groups that were unbalanced by response type. The groups of animals were subjected to an anxiolytic drug (dexmedetomidine) and compared to groups of animals injected with saline. Statistical tests were used as means to examine and interpret the results. Figure 1 gives an overview of the phases within the experiment.

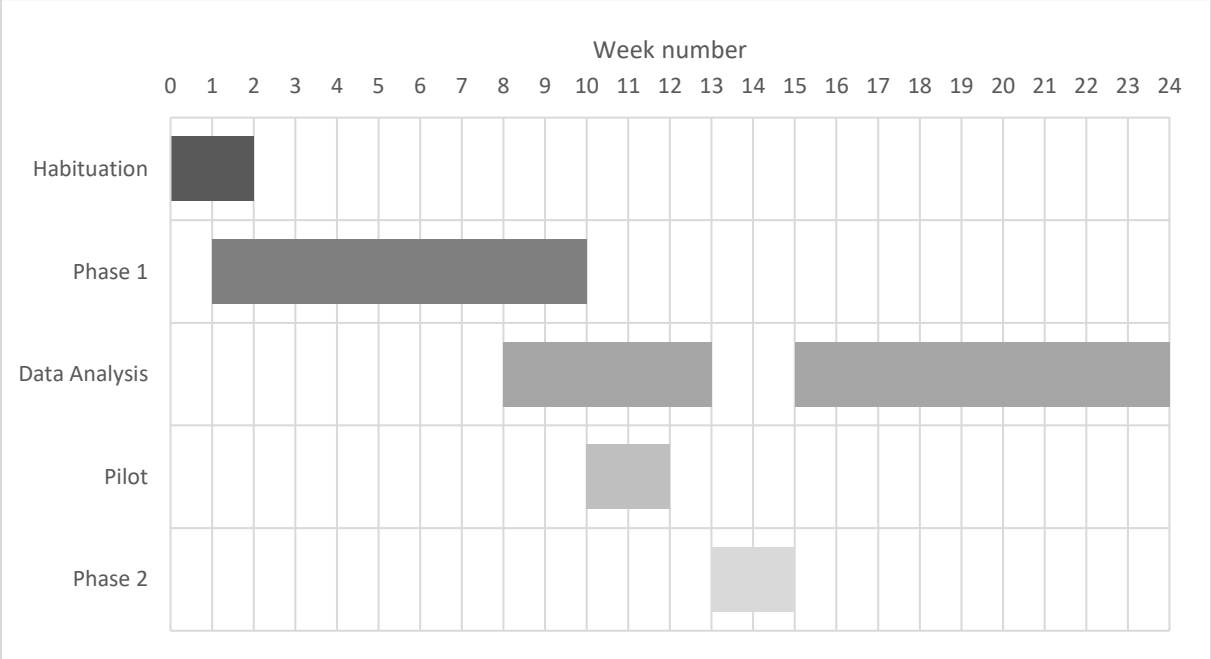


FIGURE 1. TIMELINE OVERVIEW FOR THE TOTAL EXPERIMENT. VERTICAL GRIDLINES INDICATE WEEKS.

3.3 ANIMALS AND HOUSING

In phases 1 and 2 of this experiment, *n* = 179 naïve male mice were used from the BALB/cAnNCrl (*n* = 59, albino), C57BL/6NCrl (*n* = 60, pigmented/black) and 129S2/PasCrl (*n* = 60, agouti) inbred strains. The BALB/c strain's sample size had been reduced by one mouse before the start of the experiment,

yielding a smaller sample size compared to the other two strains, due to the timely euthanasia of one of the BALB/c mice as a result of reaching the humane endpoint through health issues (50). Moreover, to establish the required dose of dexmedetomidine in a small pilot (see: 3.6 Anxiolytic drug pilot), another *n* = 15 naïve male mice were used from the same aforementioned three inbred strains (*n* = 5 per strain). During the pilot, one C57BL/6 mouse passed away as a result of experimental procedures, resulting in a C57BL/6 mice sample size of *n* = 4 in the pilot. The total amount of animals used in the present study were *n* = 193. Only male mice were used. Unneutered female mice follow an oestrus cycle that significantly alters their individual behavioural results, as a consequence of hormonal changes (51–54). All mice were seven weeks old upon arrival and they were bred by and purchased from Charles River Laboratories in the Netherlands.

The animals were all housed in Euro-standard Type II L cages (365 × 207 × 104, floor area: 530 cm^2; Techniplast, Milan, Italy) on a 12 hour reversed day-night cycle with lights on at 7 PM and off at 7 AM and a radio playing 24 hours a day. Testing was done during the dark-period of the day-night cycle, while red lights were on. The standard home cage contained aspen bedding (Aspen chips; Abedd-Dominik Mayr KEG, Köflach, Austria), food (CRM, Expanded, Special Diets Services Witham, England) and water, available ad libitum. A plastic shelter and tissue paper (KLEENEX® Facial Tissue, Kimberly-Clark Professional BV, Ede, The Netherlands) served as cage enrichment in each cage. The mice were held at the Central Laboratory Animal Research Facility of Utrecht University for the entirety of the experiment (including pre-experimental period) and were kept in two separate rooms, room K and room L. The mice were maintained under constant temperature (room K: 21.7 °C (± 0.234); room L: 21.9 °C (± 0.403)) and humidity (room K: 53.5% (± 2.44); room L: 54.8% (± 2.56)) (see appendix 2 for graphs on temperature and air humidity in both rooms during the study). Distribution of the mice over the two rooms was done at random. The animals were given two weeks to habituate prior to behavioural testing and were tested in the same room as the one in which habituation took place.

Furthermore, the test animals were being housed individually, due to the fact that the prior experiment (Exp. 1) required the socially-housed animals to be separated halfway through the experiment, because of continuous fighting between the male mice. Moreover, research indicates there are no effects of individual housing on behaviour in the mHB in mice (55), thus substantiating the protocol.

Lastly, the means of initial body weight upon arrival of the 129S2/PasCrl (129S2) mice, the BALB/cAnNCrl (BALB/c) mice and the C57BL/6NCrl (C57BL/6) mice were respectively 24.4 ± 2.09 grams, 20.7 ± 1.26 grams and 21.2 ± 1.56 grams. One outlier was not taken into account when calculating these means and standard deviations: a BALB/c mice arrived with an initial weight of 13.5 grams, showed signs of being ill and turned out to be dehydrated. The condition of this mouse had improved significantly after three days and the mouse was part of the overall experiment after recovery.

3.4 MODIFIED HOLE HOARD AND Z-SCORING

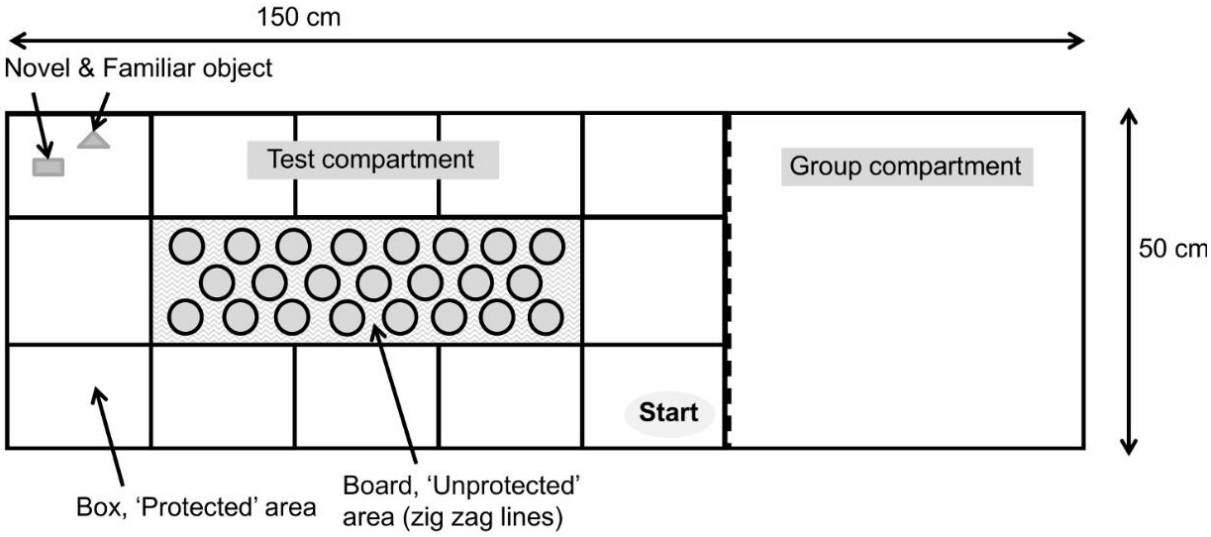


FIGURE 2. OVERVIEW OF THE MHB. "*THE SET-UP CONSISTS OF A TEST COMPARTMENT (BOX) WITH IN THE MIDDLE THE (UNPROTECTED) BOARD INDICATED WITH ZIGZAG LINES AND THE GROUP COMPARTMENT.* = *NOVEL OBJECT,* = *FAMILIAR OBJECT,* Start = *STARTING POINT.*" (28)

All behavioural tests were performed with the use of the mHB. As was mentioned before (see: 1.3 Background of the modified Hole Board), the mHB combines the characteristics of an open field test and a hole board test, resulting in a paradigm with the possibility to test a great variety of behavioural dimensions (27,28,32). Figure 2 shows a schematic top view of the board, illustrating the open field with holes in the centre of the board. None of these holes were filled. Moreover, a novel and a familiar object were placed in the box, on the opposite side of the starting point, although the data collected from interaction with these objects was not taken into account in the statistical analysis. This collected data was thought to be of use in retrospective analysis in other studies. Additionally, this experiment did not make use of the group compartment, due to the animals being individually housed. A spotlight was placed above the mHB with its light carefully positioned to only shine on the board in the centre. This was done to create an aversive effect. The mean illuminance on the board in room K was 147 (± 40.2) lx and the mean illuminance on the board in room L was 151 (± 26.1) lx. A video camera was also positioned over the test set-up to record the animals during testing, for archiving purposes.

As was previously mentioned, the five behavioural dimensions that have been scored in the present study were avoidance behaviour, risk-assessment, locomotion, exploration and arousal. These dimensions can be distributed in two different behavioural sets, namely, anxiety-related behaviour (consisting of avoidance behaviour, arousal and risk-assessment) and activity-related behaviour (consisting of exploration and locomotion). The following variables have been scored in relation to these behavioural dimensions and their behavioural sets:

I. Anxiety-related behaviour
    a. Avoidance behaviour
        i. Frequency of board visits
        ii. Latency to initial board visit
        iii. Percentage of time spent on board
    b. Arousal
        i. Grooming-bouts
            1. Frequency of grooming-bouts
            2. Latency to grooming-bouts

3. Percentage of time grooming

ii. Defecation

1. Frequency of defecation
2. Latency to defecation

c. Risk-assessment

i. Frequency of risk-assessments
ii. Latency to initial risk-assessment

II. Activity-related behaviour

a. Locomotion

i. Line crossing

1. Frequency of line crossing
2. Latency to initial line crossing

ii. Immobility

1. Frequency of immobility
2. Latency of immobility
3. Percentage of time spent on immobility

b. Exploration

i. Rearing

1. Frequency of rearing
2. Latency to rearing

ii. Exploration of test set-up

1. Frequency of exploration of holes on the board
2. Latency to exploration of holes on the board

The above-displayed behavioural variables were organised into their behavioural dimensions through z-scoring. First of all, the variables of each individual were normalised into z-scores, by using the following equation: $\text{z-score} = \frac{x-\mu_{pooled}}{\sigma}$. In this equation, x indicates an observation, σ indicates the standard deviation of said observation and μ_{pooled} describes the mean of the pooled data on this specific variable, regardless of strain and number of trials. This results in an equation that describes how many standard deviations an observation is above or below the specified mean. Even though some discrete variables were taken into account, all variables were treated as continuous to calculate the standard deviation and mean. After calculating the individual z-scores per variable, the z-scores per behavioural dimension were calculated. Appendix I gives an overview of the effect of the different variables in the five behavioural dimension. The directionality of scores was adjusted so that positively-affecting score values reflected increased dimensionality, whereas negatively-affecting score values reflected decreased dimensionality. These calculations resulted in five different z-scores per individual mouse per trial that amounted to five z-score trajectories per individual mouse.

3.5 PHASE 1 RESEARCH PROTOCOL

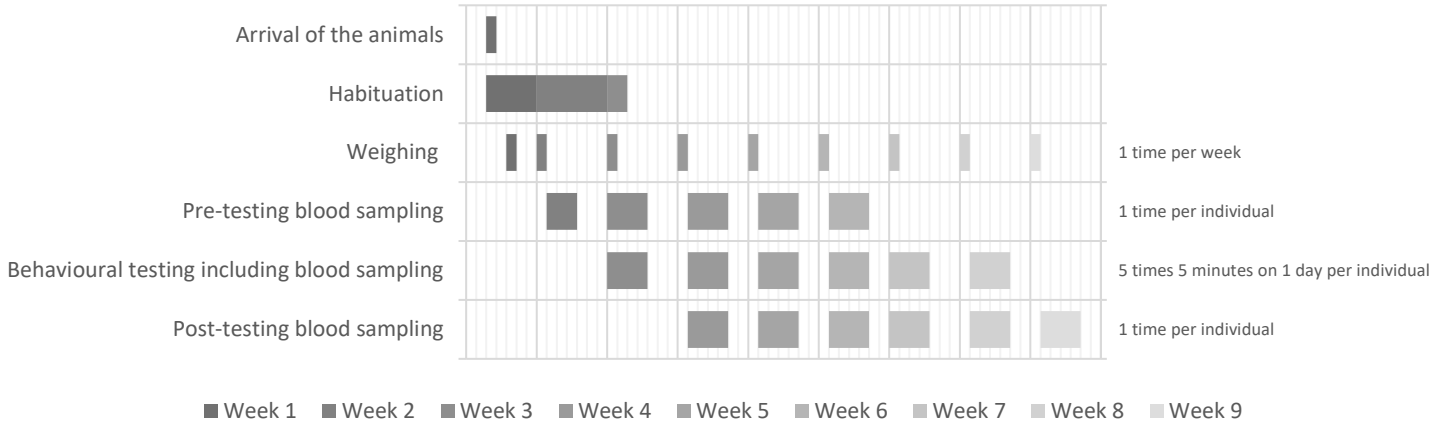


FIGURE 3. TIMELINE OF PHASE 1. THE ANIMALS ARRIVED ON WEDNESDAY THE 9TH OF JANUARY. PRIMARY VERTICAL GRID-LINES INDICATE WEEKS; SECONDARY VERTICAL GRID-LINES INDICATE DAYS. AN INDIVIDUAL MOUSE WAS SUBJECTED TO BEHAVIOURAL TESTING FIVE CONSECUTIVE TIMES ON ONE DAY.

3.5.1 BEHAVIOURAL ANALYSIS

After arriving, all *n* = 179 animals were randomly divided into two rooms (*n* = 89 in room K, *n* = 90 in room L), with equal distribution of the strains. Figure 3 shows a schematic timeline of phase 1 starting with the arrival of the animals. Before the start of the behavioural analyses, all animals were allowed acclimatisation for at least two weeks. During these two weeks, the animals were handled three times per week and weighed at least one time a week. Handling was done to habituate the mice to the researchers and to handling, which would consequently decrease the levels of stress associated with handling (56,57). During weighing and handling, general impression, including behaviour and body condition score (58–60), was also examined.

Following the first two weeks, behavioural testing started. Two experimenters had been trained together to score and observe the behavioural parameters in the mHB and their results were well-matched with inter-rater reliability results of over 80%. Each observer had their "own" room with mice to observe. At the start of behavioural testing, an individual mouse would be placed in the mHB compartment and was allowed free exploration for five minutes, while the observer scored the behaviour using the computer software Observer XT 12.5 (Noldus, Wageningen, The Netherlands) (61). After mice had been subjected to the mHB for five minutes, they were placed back into their home cage and the mHB was cleaned with a damp paper towel before the next trial started. This was repeated five consecutive times per individual mouse on a single day between 10 AM and 2 PM. Between the tests, the mHB was cleaned and wiped with paper towels and water. After all tests for the day were completed, the mHB was cleaned with alcohol as well. Testing blindly on strain was not possible due to differences in colour of the coats.

3.5.2 BLOOD SAMPLING AND CORTICOSTERONE LEVELS

A total of three blood samples were collected per individual animal. The first blood sample was taken one week (approximately seven days, with a minimum of six days and a maximum of eight days) before the start of the behavioural tests to establish a base level of corticosterone per individual. Next, the second blood sample was taken as soon as possible after the fifth and final behavioural test. Finally, the third and last blood sample was taken one week (approximately seven days, with a minimum of six days and a maximum of eight days) after the behavioural tests had taken place. The individual mice

always had their three blood samples taken at approximately the same time on each sampling day, since corticosterone has been known to have a circadian rhythm (62).

The samples were all taken in a different room than where the animals were housed. The animals were transported to this room while being situated in the home cage and by covering the home cage with a towel, after which the cage was carried by one of the observers to the room used for taking blood samples. The time from the moment the home cage was lifted from its usual spot until the moment enough blood was collected was noted.

All samples were collected by making a tail incision with a single edge industrial blade (GEM®: SPI Supplies, West Chester, PA, USA) at a 90° angle in respect to the long axis of the tail along the underside of the tail. The tail incision for the first blood collection was made farthest from the base of the tail compared to the other two tail incisions, whereas the tail incision for the third blood collection was made closest to the base of the tail. The blood drops were collected in pre-chilled EDTA coated Microvette® CB300 capillaries (Sarstedt, Nümbrecht, Germany). After blood collection, the animals were held outside the cage to observe and compress the bleeding when necessary, until the bleeding had stopped sufficiently and they were placed back into their home cage. The capillaries were stored on ice and plasma was collected by centrifuging the capillary tubes for 30 minutes at 3000 rpm (diameter of the rotor: 17 cm) in a refrigerated centrifuge (IEC Microlite/Microlite RF®: Thermo Electron Cooperation; West Sussex, UK) that was set at 4 °C. After centrifuging, 10-20 µL plasma was pipetted into microtubes. If the plasma seemed to have a red hue or if it was mixed with blood by accident, it was centrifuged again at the same settings for 5 minutes. If plasma retained a pink or red hue even after centrifuging a second time, haemolysis was noted down for that specific sample and the plasma was still collected and stored. The collected plasma was stored at -26 °C.

The blood plasma corticosterone levels were analysed with the use of use of the Corticosterone Double Antibody RIA Kit (MP Biomedicals, LLC, Orangeburg, NY, USA). This RIA kit (Lot No. RCBK1905,06) was utilised according to the supplied manufacturer's protocol and testing was done in duplicate. Randomisation was used on the order in which the blood plasma corticosterone levels in the samples were determined, although duplicate samples and samples of the same animal were kept together during testing. Quality control was to be secured through the use of a high and low control, which were determined beforehand by the manufacturer to be within a certain range. The results of the RIA in counts per minute were transformed into concentration units with the use of calibration curves.

However, since the duplicate samples were sometimes not in correspondence with each other, an attempt was made to retest the samples that showed a difference in the coefficient of variation of over 50% between their duplicates. Unfortunately, because the rerun showed higher average results than the original results, the new results could not be used. Therefore, it was decided to correct the sample results if the average of an original pair of duplicates was lower than the average of the same sample in the rerun. If this criterium was met, the original duplicate sample with the largest deviation of the retested sample average would be discarded.

3.5.3 Statistical analysis

Analysis on the collected data was done with the use of R 3.5.1 and RStudio (63,64). The R-packages cluster (65), nlme (66), glmmTMB (67) and kml3d (25) were also utilised in the analysis.

The linear mixed models were constructed per behavioural dimension z-score to create residuals that could be used in the longitudinal k-means clustering analyses. Linear mixed models usually consist of fixed effects that are being tested for influencing a dependent variable. However, to create models that were fit for cluster analysis, fixed effects were merely used to control potentially confounding

factors. Experimenter (i.e. a factor representing the observer doing the behavioural testing and the specific room an individual mouse was housed in) and strain were used as fixed effects and a list of random effects was used in the following order: test group, test order and mouse ID. While experimenter, strain, test group and mouse ID consisted of factorial data, the z-scores and test order contained continuous data. Test group was defined by the week in which the mice were tested and was consequently a proxy for the inter-individual difference in age. Distribution of the residuals and homogeneity of variances of the models were assessed with the use of a Shapiro-Wilk Normality test, a histogram, a QQ-plot and a residual plot. The avoidance model was log transformed, the arousal model was rank transformed, the locomotion model was square root transformed and the exploration model was rank transformed. A variance transformation for intra-strain variances was applied in all four aforementioned models.

However, linear mixed models on risk-assessment with normal distribution could not be produced in the same way: due to zero-inflated effects and the small number of behavioural variables that the dimension consisted of, a zero-inflated mixed model was used (67). Firstly, all dimensional z-scores that were the equivalent to a score of zero, were substituted by literal zeroes. The model was then fitted similarly to the aforementioned models with experimenter and strain as fixed effects and test group, test order and mouse ID as random effects. The one-sided zero-inflated formula was decided to consist of experimenter and strain and a gamma distribution was used. The residuals of this model were checked on normal distribution with the use of a Shapiro-Wilk Normality test, a histogram and a QQ-plot and were accepted on the combined results of these tests. It was not possible to take strain into account as an effect on variance in response, due to the model being zero-inflated.

Moreover, regarding the corticosterone data, individuals that were missing two or more blood samples and/or the blood sample that was taken after behavioural testing were excluded from the data analysis on clustering, since the trajectories per individual mouse could not yield reliable results. The curve that was expected could not be created correctly when said blood samples were missing, possibly influencing the cluster an individual was put in majorly. The corticosterone data was also transformed into z-scores and was fitted into a linear mixed model using experimenter and strain as fixed effects and test group, test order and mouse ID as random effects. A log transformation was used on the data to create normally distributed residuals and strain was used to describe the within-group heteroscedasticity structure.

The residuals of all six aforementioned models were put into one dataset and were standardised using z-scoring per dimension per trial. K-means clustering for longitudinal data was utilised on this dataset. This meant that the z-scores per trial and per behavioural dimension could be inserted into the analysis as joint trajectories for each individual mouse. The dataset was searched on partitions with 2 to 6 clusters. K-means were run 100 times on each cluster. The Calinski-Harabasz, Ray-Turi and Davies-Bouldin criteria were used as the criteria to determine the most optimal number of clusters (25). Dimensional reduction was utilised if chi-squared tests showed no difference between including or excluding a certain dimension (68). As a result, the corticosterone data and the behavioural dimensions arousal and risk-assessment were excluded from the cluster analysis.

Nonetheless, corticosterone levels were still tested on influence of strain and cluster with the use of linear mixed models, to investigate inter-strain differences and inter-cluster differences in corticosterone levels. No individuals and/or samples were left out this time, as these models would not yield results per individual, in contrast to the linear mixed models on corticosterone data used in the cluster analysis. In the first model, strain, trial, experimenter and all possible interactions between these three variables were used as fixed factors and test group and mouse ID were used as random factors, with test group nested in mouse ID. A second, similar model was created using cluster instead

of strain as one of the three main factors. Model reduction by least-significant difference and single term deletions based on the Akaike information criterion were utilised to simplify the models, although none of the main effects were to be removed from the models. The models were both transformed using rank transformation. The standardised Pearson residuals were tested on normal-distribution with the use of a Q-Q plot, a histogram and a Shapiro-Wilk Normality Test. If significant p-values were found in one or several factors in the models, planned comparison of means on the transformed model was done by using a pairwise comparison of contrasts. Not all contrasts were included, as not all contrasts were considered relevant. To elucidate, strains were only compared to each other per sampling moment and sampling moments were only compared per strain. No comparison between strains on different sampling moments was done. A similar method was used on the cluster-effect-models.

3.6 ANXIOLYTIC DRUG PILOT PROTOCOL

3.6.1 ANIMALS AND HOUSING

n = 14 naïve male mice from the BALB/cAnNCrl (*n* = 5, albino), C57BL/6NCrl (*n* = 4, pigmented/black) and 129S2/PasCrl (*n* = 5, agouti) inbred strains were used in the pilot of this experiment. The animals were housed in exactly the same way as the *n* = 179 animals of phase 1 and 2 and were kept in the same two rooms as the *n* = 179 animals of the main experiment, as to ensure the standardisation in environment of the *n* = 179 mice. The animals arrived when they were seven weeks of age and were used in the pilot study after two initial weeks of habituation, when they were nine and ten weeks of age.

3.6.2 RESEARCH QUESTION

The anxiolytic drug pilot was used as a means to establish the required dose in the three distinctive strains. The required dose was defined to 1) increase exploratory behaviour; 2) decrease risk-assessment and avoidance behaviour; 3) not influence locomotion and 4) not produce sedation. The anxiolytic drug that was used was dexmedetomidine (Dexdomitor®, Orion Corporation–Orion Pharma, Espoo, Finland) (see box 1).

Box 1. Dexmedetomidine dose in the 129S2, BALB/c and C57BL/6 strains

Dexmedetomidine is a highly selective α_2-adrenergic receptor agonist (69). The dose of dexmedetomidine as an anxiolytic had only previously been established in the C57BL/6 mice (70). However, this study involved a different substrain than used in the present study and the implemented dose did not affect avoidance behaviour (70).

The required dose of dexmedetomidine needed to be established in all three mouse strains. It was decided that this dose had to be the same in all three inbred strains to ensure standardisation. To test the influence of the dose of dexmedetomidine, the mice that were used in the pilot were tested in the mHB, using the same ethogram as in phase 1 and phase 2 of this study (see: 3.4 Modified Hole Board and z-scoring). In advance, solutions with six different concentrations (2.5, 5.0, 7.5, 10.0, 12.5 and 15.0 µg/kg body weight) were prepared with the use of saline (NaCl 0.9%, B. Braun, Melsungen, Germany). The order of injection and behavioural testing was randomized per strain and three mice of the three different strains were tested consecutively. Mice were weighed several minutes before the injection.

3.6.3 RESEARCH PROTOCOL

Determining the required dose was done on two different days between 11 AM and 3 PM. At first, three mice of the three distinctive strains were weighed and then injected intra-peritoneally with 10.0 µg/kg dexmedetomidine solution by the experimenter who would not do the behavioural testing on that individual animal. Behavioural testing started approximately 30 minutes after injection. After

behavioural testing of the three animals, their results on latency and duration of board visits, amount and latency of risk-assessments, amount and latency of line crossings and amount and latency of hole explorations were compared to the average and median of those behaviours in the first trial of their respective strains during phase 1. The next concentration of injection was decided based on this comparison, at first using steps to decrease or increase the dose with 5.0 µg/kg, meaning that the next dose would be either 5.0 µg/kg or 15.0 µg/kg; three other animals were injected with the decided doses and were tested similarly to the first three animals. If deemed necessary on the basis of the previous results, a next dose could use a step of 2.5 µg/kg, leading to a dose of either 7.5 µg/kg or 12.5 µg/kg. This step was only regarded as necessary in the 129S2 strain. In the other two strain, only doses of 5.0 µg/kg, 10.0 µg/kg and 15.0 µg/kg were tested.

Subsequently, the dose of 10.0 µg/kg was deemed most ideal considering all three strains. To validate this decision and create more accurate results, two more mice of both the 129S2 and BALB/c strains and one more mouse of the C57BL/6 strain were injected intra-peritoneally with 10.0 µg/kg dexmedetomidine solution and tested in the mHB.

3.6.4 Statistical analysis

Analysis on the collected pilot data was only done on the basis of visual parameters (see: 4.2 Anxiolytic drug pilot). It was decided not to run any statistics on this data, due to the limited sample size. Any result that might have been obtained would consequently lose any reliability.

3.7 PHASE 2 RESEARCH PROTOCOL

3.7.1 Test groups

To accurately test the influence of the treatment on behaviour, mice were to be put in pairs to make more reliable comparisons between the treatments. Consequently, with the use of k-means clustering on all five behavioural dimensions, it was tested which individuals per strain and per cluster complemented each other the most and sub-clusters were created. K-means cluster analysis that worked on joint trajectories were run 100 times on each sub-cluster and the partition was chosen to be half of the tested subset, rounded off in case the subset consisted of an odd number of animals. Next, it was attempted to create the pairs by using experimenter, sub-cluster and last measured individual body weight as factors that majorly contributed to the resemblance between the animals. On the basis of this resemblance, different mice were paired off and placed into a group balanced in terms of response cluster, whilst other mice showed weaker relation to others and were placed in another group which was unbalanced in terms of response cluster. Both the balanced and unbalanced groups were composed to be of equal size (n = 82 per group). The pairs in the unbalanced group were ultimately created using only body weight and experimenter as factors. This meant that each pair consisted of two mice of the same strain with resembling weights and, if part of the balanced group, additionally showed similar z-scoring in the five behavioural dimensions considering the five consecutive trials of phase 1. Within each pair, one mouse was injected intra-peritoneally with 10.0 µg/kg dexmedetomidine (Dexdomitor®, Orion Corporation–Orion Pharma, Espoo, Finland) and one mouse was injected intra-peritoneally with saline (NaCl 0.9%, B. Braun, Melsungen, Germany). This was done at random per pair and the mouse injected with saline received the same injection volume in relation to its body weight as the mouse injected with dexmedetomidine. Finally, the balanced pairs were matched between the two different rooms and put into different blocks that aimed to include all strains and all clusters. This is illustrated in figure 4 and the result was four different blocks. It should be noted that not all blocks could be created in the same way, due to unequal distribution of clusters among strains and rooms. The aim, however, was to pursue this construction per block to the largest extent possible. The

unbalanced pairs were only matched between the two different rooms and thus were not described in different blocks.

However, during the experiment, one of the animals passed away before its behavioural testing was done as a result of the experimental procedure. Therefore, it was decided to exclude the animal with which the deceased mouse had formed a pair. Additionally, it was decided to exclude a pair of the same strain in the unbalanced group that was tested on the same day, in the same room and nearest to the time the balanced pair was to be tested. This decision was made on the basis that different mixed models would be made on the data of balanced and unbalanced groups. Comparison between the models of the balanced and unbalanced group could only be executed accurately if the sample size in both models was equal. As a result, the sample sizes of both the balanced and unbalanced group were reduced by two mice each, meaning that a total of four mice were removed from the experiment. The final sample sizes in the balanced and unbalanced groups were n = 80 per group.

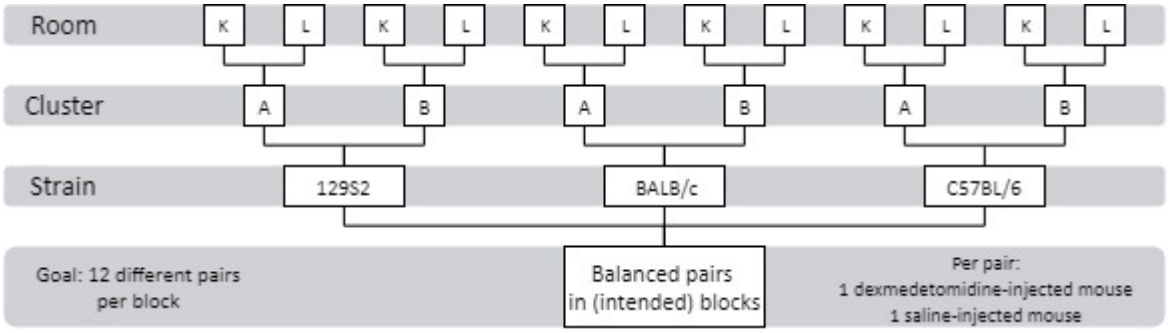


FIGURE 4. DIAGRAM OF INTENDED CONSTRUCTION OF BLOCKS. TREATMENT IS NOT INCLUDED IN THE DIAGRAM.

3.7.2 BEHAVIOURAL ANALYSIS

Behavioural testing was done on four consecutive weekdays with a weekend in-between, starting on a Thursday and ending on a Tuesday. With a set-up similar to phase 1 (see: 3.5.1 Behavioural analysis), the same two experimenters did behavioural testing in the two rooms where the animals were kept in. Both experimenters did observations in the same room as they did observations in during phase 1. The individuals were still housed in the same manner and in the same room as during phase 1. The four blocks were all tested on separate days. Balanced and unbalanced pairs were alternately tested per day; paired animals of the same quartet in both the balanced and unbalanced group (see: 3.7.1 Test groups) were tested at the same time in the two different rooms. The mHB was used again and all behavioural tests were run between 9.30AM and 2.30PM. Between 9AM and 9.30AM all mice were weighed to determine the injection volume. All animals were injected 30 minutes before behavioural testing by a third researcher who did not do behavioural observations in this experiment. The experimenters that were doing the behavioural testing were blinded regarding the treatment, the pairs and groups of the animals (balanced vs. unbalanced). All injections were done intra-peritoneally and in the same room as in which the animals were housed. After a test, the mouse was placed back into its home cage and the mHB was cleaned and wiped with paper towels and water. After all tests for the day were completed, the mHB was cleaned with alcohol as well. Behaviour was scored with the same ethogram as before (see: 3.4 Modified Hole Board and z-scoring) and with the use of the computer software Observer XT 12.5 (Noldus, Wageningen, The Netherlands) (61) on a laptop.

3.7.3 STATISTICAL ANALYSIS

Analysis on the data collected in phase 2 was again done by utilising R 3.5.1 and RStudio (63,64) and the R-packages nlme (66), emmeans (71), car (72) and ggplot2 (73). Transformation of raw data to integrated behavioural z-scores was done on the balanced and unbalanced group separately and in

the same way as described in the case of phase 1 (see: 3.4 Modified Hole Board and z-scoring). All five dimensions were taken into account. Linear mixed models were used to analyse correlations between different factors and the five behavioural dimensions, resulting in 10 different models: the balanced and unbalanced group, both divided into five different behavioural dimensions. The models fitted experimenter, strain, treatment and all possible interactions as fixed factors and fitted ID's for pairs nested in test day as random factor. To simplify the models, model reduction by least-significant difference and single term deletions based on the Akaike information criterion (AIC) were used, the non-interactive, main factors were never excluded. The standardised Pearson residuals were tested on being normally distributed with the use of a Shapiro-Wilk Normality Test, a QQ-plot and a histogram. Transformations were done after simplifying the models to achieve normal distribution in the residuals, although the two models of the same dimension (balanced and unbalanced) were always transformed using the same method. This resulted in the avoidance behaviour models being rank-transformed and the arousal and exploration models being subjected to square root transformation. The other four models of risk-assessment and exploration were not transformed. Variance in the residuals was tested with a scatterplot and Levene's test and was found to be homoscedastic in all models.

The significance level was determined as $\alpha = 0.05$. The p-values of the transformed models were obtained through Chi-square tests, as provided by the R-package car (72). Type II was used when no interactions were present in the final model, whereas type III was used when interactions were present in the final model. In case of a significant fixed factor, planned comparison testing was done on the transformed models by utilising a pairwise comparison of contrasts. Dunn–Šidák correction was used as a means to correct alpha in the paired comparisons (74). Moreover, as the residuals' variance of all 10 models were homoscedastic, it was possible to execute F-tests to compare the variances in residuals between the balanced and unbalanced groups per behavioural dimension.

4. RESULTS

4.1 PHASE 1

Figure 5 presents the final clusters as line charts. After visualisation of the clusters, the corticosterone data and the behavioural dimensions arousal and risk-assessment did not seem to differ much per cluster. With the use of chi-squared testing, it was found that there was no significant difference in cluster membership of individual animals when including or excluding corticosterone data and these two dimensions (χ^2 (1) = 0.45, p = 0.916). To simplify the data, it was thus decided to exclude all three in the cluster analysis. The remaining dimensions avoidance, exploration and locomotion were thus used to characterize mice on their response type. The graphs presented in figure 5 show the two discarded behavioural dimensions, but these dimensions were not taken into account when creating the clusters. The final clustering procedure yielded two response type clusters.

The distribution of the inbred mouse strains amongst the two clusters is shown in table 1. The 129S2 mice were largely allocated to cluster A (88.3%), whereas the BALB/c mice were largely assigned to cluster B (87.7%). The C57BL/6 mice were more or less equally distributed amongst the two clusters.

TABLE 1. OVERVIEW OF DISTRIBUTION OF THE THREE DIFFERENT MOUSE STRAINS AMONGST THE TWO CLUSTERS. OVERALL DISTRIBUTION OF SAMPLE SIZE AMONGST THE TWO DIFFERENT CLUSTERS IN PERCENTAGE IS SHOWN IN THE LAST COLUMN, WHEREAS THE DISTRIBUTION OF STRAIN PER CLUSTER IN PERCENTAGE IS SHOWN BEHIND THE SAMPLE SIZE OF STRAIN PER CLUSTER.

Cluster	**129S2** (*n* = 60)	**BALB/c** (*n* = 59)	**C57BL/6** (*n* = 60)	(*n* = 179)
A	53 *(88.3%)*	9 *(15.3%)*	32 *(53.3%)*	*94 (52.5%)*
B	7 *(11.7%)*	50 *(84.7%)*	28 *(46.7%)*	*85 (47.5%)*

Cluster A is shown to increase avoidance behaviour per trial, whereas exploration and locomotion decrease per trial. Cluster B can be seen to show the opposite results, presenting an increase in exploration and locomotion, whilst avoidance behaviour decreases over time.

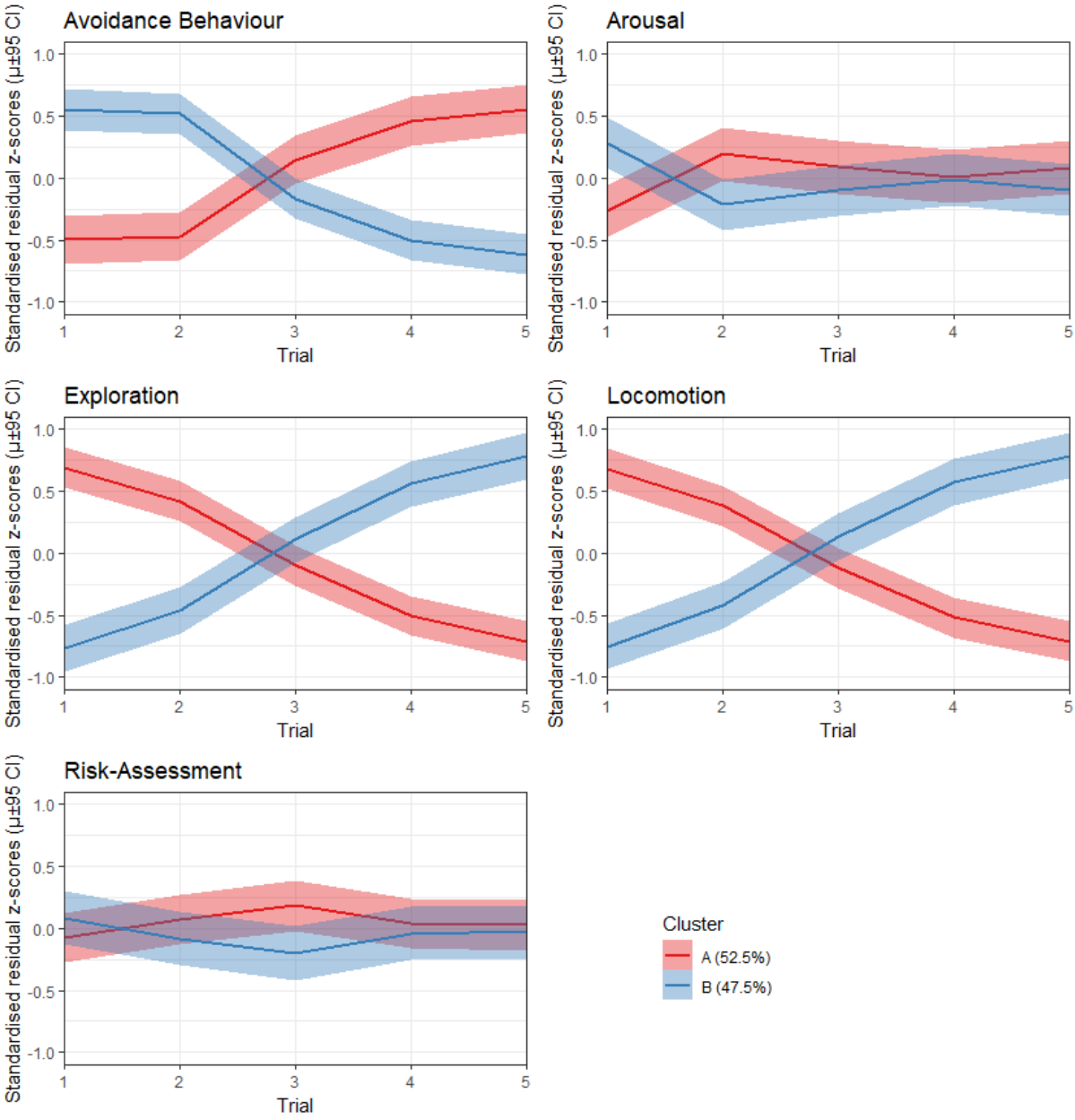


FIGURE 5. LINE CHARTS AS A RESULT OF CLUSTER ANALYSIS IN PHASE 1. THE X-AXIS PRESENTS THE NUMBER OF TRIALS, WHEREAS THE Y-AXIS PRESENTS THE STANDARDISED RESIDUALS OF THE Z-SCORES FOR THAT SPECIFIC BEHAVIOURAL DIMENSIONS. THE LEGEND SHOWS THE DISTRIBUTION OF THE SAMPLE SIZE AMONGST THE TWO DIFFERENT CLUSTERS. THE LINES SHOW THE MEAN STANDARDISED RESIDUALS OF THE Z-SCORES AND A 95% CONFIDENCE INTERVAL IS INCLUDED.

TABLE 2. TOTAL AMOUNT OF CORTICOSTERONE SAMPLES PER STRAIN (*N* = 179)

Sampling moment	129S2	BALB/c	C57BL/6
1	57	58	60
2	57	58	56
3	58	56	55

Regarding corticosterone, mean ng/mL corticosterone per strain per sampling moment are described in figure 6A. Regarding the different strains, the 129S2 and C57BL/6 strain consisted of *n* = 60 animals and the BALB/c strain consisted of *n* = 59 animals. Even though all the animals had blood samples taken three times, some samples were missing, usually because of vasoconstriction in the tail, resulting in a sample volume that was too small to use. Consequently, the total amount of samples per strain per sampling moment can be found in table 2.

TABLE 3. TOTAL AMOUNT OF CORTICOSTERONE SAMPLES PER CLUSTER (*N* = 179)

Sampling moment	Cluster A	Cluster B
1	89	81
2	85	82
3	86	79

Moreover, mean ng/mL corticosterone levels per cluster per sampling moment are described in figure 6B. An overview of the total amount of samples per cluster can be found in table 3. To clarify, both tables 2 and 3 and figures 6A and 6B make use of the same corticosterone data, but use different means to analyse the data: table 2 and figure 6A show the data of corticosterone samples per strain, whereas table 3 and figure 6B show the data of corticosterone samples per cluster.

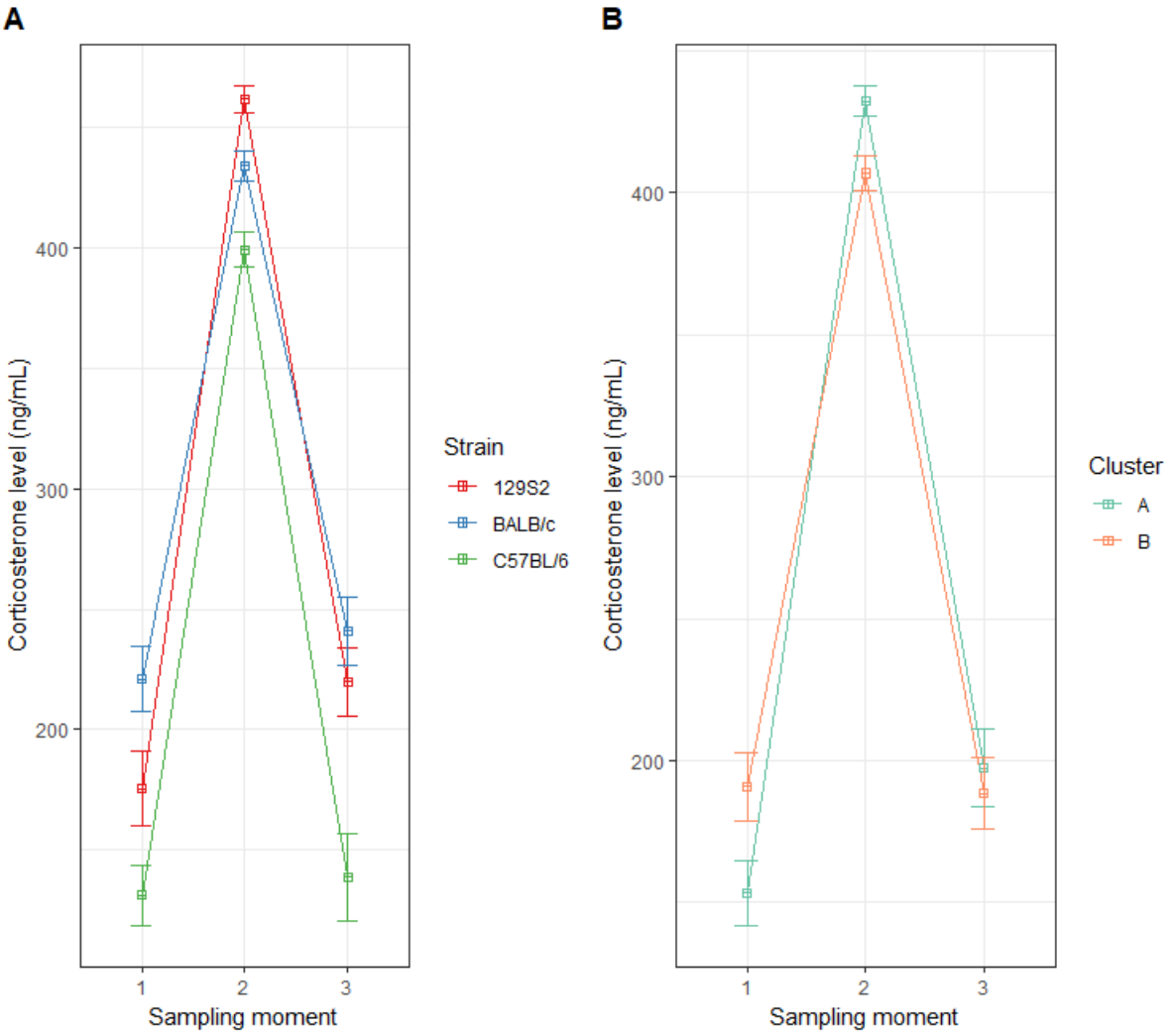


FIGURE 6. LINE GRAPHS SHOWING (RANK-TRANSFORMED) CORTICOSTERONE LEVEL IN NANOGRAM PER ML. CORTICOSTERONE LEVELS ARE PRESENTED PER TRIAL PER STRAIN (A) AND PER TRIAL PER CLUSTER (B). THE TIME BETWEEN THE SAMPLING MOMENTS WAS ONE WEEK (I.E. SEVEN DAYS). STANDARD DEVIATION PER SAMPLING MOMENT PER STRAIN IS ALSO SHOWN. THE DATA SHOWN IS ON NON-TRANSFORMED DATA.

The fitted linear mixed models on corticosterone levels showed significant results in the strain-sampling moment interaction (χ^2 (4) = 37.56, $p < 0.0001$). The results of the planned comparison on the transformed model are shown in table 4. Within the first sampling moment, the BALB/c strain showed significantly higher corticosterone levels compared to the 129S2 and C57BL/6 strains (respectively: |32.09| ± 19.0 (non-transformed), $p = 0.0019$; |37.03| ± 18.8 (non-transformed), $p < 0.0001$). Within the second sampling moment, a significant difference in corticosterone levels was only found between the 129S2 mice and the C57BL/6 mice, wherein the 129S2 mice showed higher corticosterone levels (|222.30| ± 19.2 (non-transformed), $p = 0.0001$). Finally, within the third sampling moment, C57BL/6 animals showed lower levels compared to the 129S2 and BALB/c animals (respectively: |44.87| ± 19.2 (non-transformed), $p < 0.0001$; |51.93| ± 19.4 (non-transformed), $p < 0.0001$).

Moreover, each strain showed significant differences between the first and second sampling moment and between the second and third sampling moment ($p < 0.0001$ in all six comparisons, see table 4 for non-transformed estimates and standard errors). These results demonstrated that the second

sampling moment showed significantly higher levels of corticosterone when compared to the first or third sampling moment. However, such a significant difference when comparing the first and third sampling moment was only found in the 129S2 strain (|27.25| ± 18.8 (non-transformed), p = 0.00021), which indicated that corticosterone levels from the third sampling moment were significantly higher than the corticosterone levels from the first sampling moment in this specific strain. The other two strains did not show significant differences in corticosterone levels between the first and third sampling moment.

Similar results were found in the model using cluster instead of strain: this model showed a significant cluster-sampling moment interaction (χ^2 (2) = 17.56, p = 0.0002). As shown in table 5, a significant difference in corticosterone level between clusters A and B was exclusively found in samples of the first sampling moment (|31.70| ± 17.9 (non-transformed), p = 0.0061), in which cluster B showed higher levels of corticosterone. The other sampling moments did not present such a significant difference between the two clusters. Furthermore, coinciding with the strain-sampling moment interaction, both clusters demonstrated significant results when comparing the first to the second sampling moment and when comparing the second to the third sampling moment (p < 0.0001 in all four comparison, see table 5 for non-transformed estimates and standard errors). Cluster A also showed significantly higher corticosterone levels between the first and third sampling moment (|20.25| ± 16.6 (non-transformed), p = 0.0243), whereas cluster B did not present complementary significance between these two sampling moments.

TABLE 4. PLANNED COMPARISON ON STRAIN-SAMPLING MOMENT INTERACTION. "-" INDICATES A BETWEEN-COMPARISON BETWEEN TWO GROUPS. THE FIRST 9 ROWS PRESENT COMPARISON PER SAMPLING MOMENT BETWEEN STRAINS, WHEREAS THE LAST 9 ROWS PRESENT COMPARISON PER STRAIN BETWEEN SAMPLING MOMENTS. THE PRESENTED ESTIMATE AND STANDARD ERROR (SE) ARE DESCRIBED IN NON-TRANSFORMED FORMAT. "*" INDICATES A SIGNIFICANT RESULT IN THE COMPARISON AFTER ALPHA-CORRECTION (P < 0.05).

Sampling moment	*Strain*	Estimate	SE	P-value
1	*129S2 - BALB/c*	-32.09	19.0	*0.0019
1	*129S2 - C57BL/6*	4.94	18.9	0.7681
1	*BALB/c - C57BL/6*	37.03	18.8	*<0.0001
2	*129S2 - BALB/c*	87.46	19.0	0.701
2	*129S2 - C57BL/6*	222.30	19.2	*0.0001
2	*BALB/c - C57BL/6*	134.84	19.1	0.0760
3	*129S2 - BALB/c*	-7.06	19.1	0.9978
3	*129S2 - C57BL/6*	44.87	19.2	*<0.0001
3	*BALB/c - C57BL/6*	51.93	19.4	*<0.0001
1-2	*129S2*	-482.33	18.9	*<0.0001
1-3	*129S2*	-27.25	18.8	*0.0021
2-3	*129S2*	455.08	18.8	*<0.0001
1-2	*BALB/c*	-362.79	18.7	*<0.0001
1-3	*BALB/c*	-2.23	18.9	1.0000
2-3	*BALB/c*	360.56	18.9	*<0.0001
1-2	*C57BL/6*	-264.97	18.8	*<0.0001
1-3	*C57BL/6*	12.68	18.9	0.5364
2-3	*C57BL/6*	277.65	19.2	*<0.0001

TABLE 5. PLANNED COMPARISON ON CLUSTER-SAMPLING MOMENT INTERACTION. "-" INDICATES A BETWEEN-COMPARISON BETWEEN TWO GROUPS. THE FIRST 3 ROWS PRESENT COMPARISON PER SAMPLING MOMENT BETWEEN STRAINS, WHEREAS THE LAST 6 ROWS PRESENT COMPARISON PER STRAIN BETWEEN SAMPLING MOMENTS. THE PRESENTED ESTIMATE AND STANDARD ERROR (SE) ARE DESCRIBED IN NON-TRANSFORMED FORMAT. "*" INDICATES A SIGNIFICANT RESULT IN THE COMPARISON AFTER ALPHA-CORRECTION ($P < 0.05$).

Sampling moment	*Cluster*	Estimate	SE	P-value
1	*A-B*	-31.70	17.9	*0.0061
2	*A-B*	64.25	18.1	0.8186
3	*A-B*	-1.24	18.2	1.0000
1-2	*A*	-416.66	16.6	*<0.0001
1-3	*A*	-20.25	16.6	*0.0243
2-3	*A*	396.41	16.8	*<0.0001
1-2	*B*	-320.70	17.1	*<0.0001
1-3	*B*	10.21	17.3	0.9691
2-3	*B*	330.91	17.2	*<0.0001

4.2 ANXIOLYTIC DRUG PILOT

An overview of the anxiolytic drug pilot results can be found in figure 7. The 129S2 strain, the BALB/c strain and the C57BL/6 are represented each in groups of 9 graphs (respectively A, B and C). Note that not all strains had been tested with the same three doses. As a result, the x-axis of the graphs may differ per strain. The graphs indicate that locomotor activity (in this pilot represented by the number of line crossings and latency until first line crossing) was severely affected in all three strains when the utilised dose was 15.0 µg/kg. When looking at graph-group C in figure 7, it shows $n = 2$ regarding a dose of 10 µg/kg, whereas the other strains show $n = 3$ with that same dose. This is due to the timely death of one mouse, because of a bladder rupture after intra-peritoneal injection.

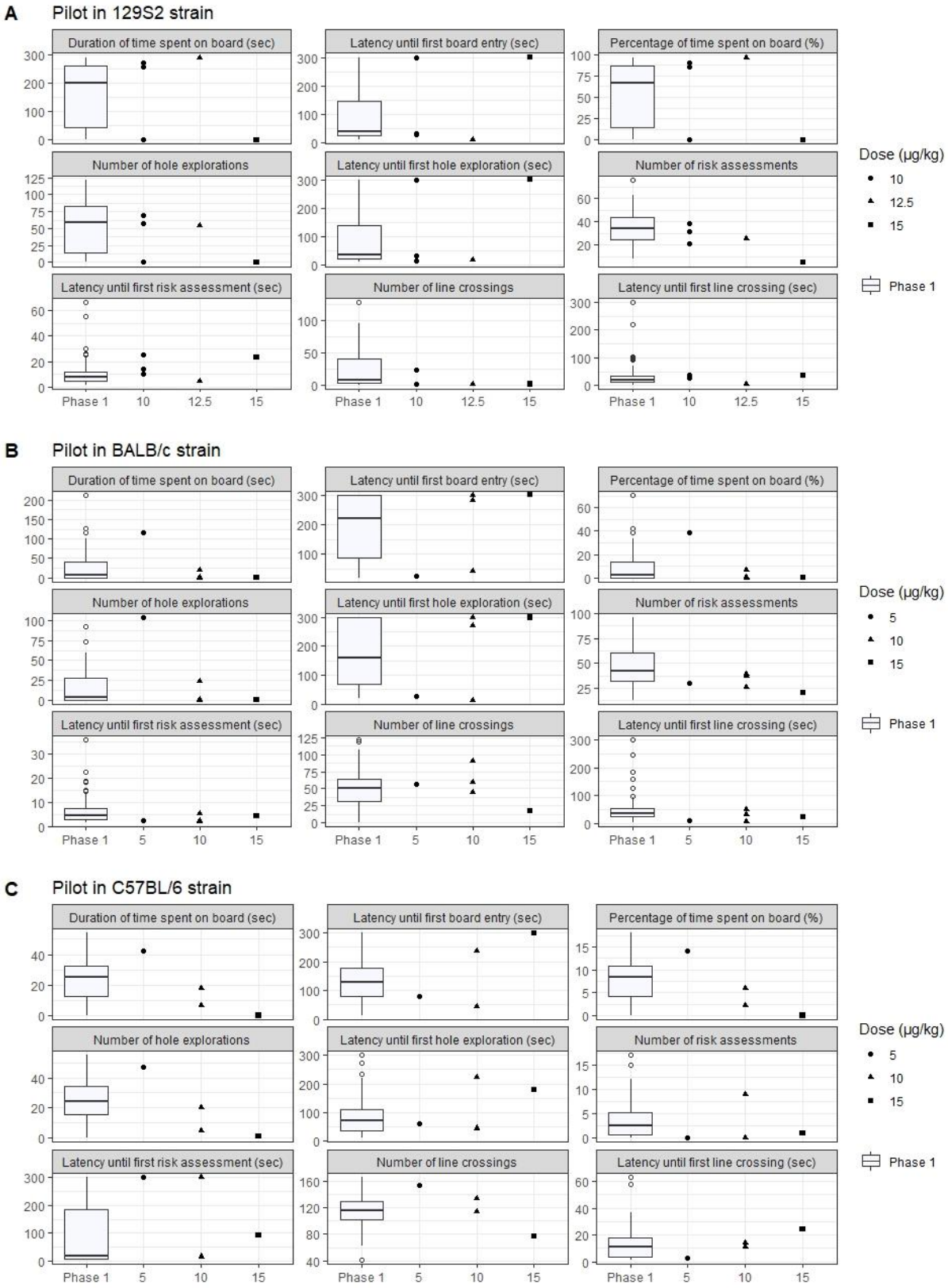


FIGURE 7. OVERVIEW OF ANXIOLYTIC PILOT RESULTS. THE GRAPHS IN "A" SHOW THE RESULTS IN THE 129S2 STRAIN, THE GRAPHS IN "B" SHOW THE RESULTS IN THE BALB/C STRAIN AND THE GRAPHS IN "C" SHOW THE RESULTS IN THE C57BL/6 STRAIN. EVERY GRAPH REPRESENTS A BEHAVIOURAL VARIABLE PER STRAIN. THE BOXPLOTS IN EACH GRAPH SHOW THE RESULTS OF THAT SAME IN BEHAVIOUR WITHIN THAT STRAIN DURING THE FIRST TRIAL OF PHASE 1. NOTE THAT EACH GRAPH USES DIFFERENT Y-AXIS LIMITS.

4.3 PHASE 2

The sample sizes that were used in phase 2 in the balanced and unbalanced groups were both *n* = 80 (see: 3.7.1 Test groups). This can be seen in tables 6 and 7. As stated before, due to the death of one individual, two animals from each group – a total of four animals – were excluded from the analysis (see: 3.7.1 Test groups).

TABLE 6. OVERVIEW OF SAMPLE SIZE OF THE GROUP OF ANIMALS THAT WAS BALANCED IN TERMS OF RESPONSE TYPE CLUSTER. "(-N)" INDICATES THE DECISION TO EXCLUDE N ANIMAL(S) IN THIS GROUP/SUBSET AFTER FINISHING THE PRACTICAL EXPERIMENT AND DATA COLLECTION, BUT BEFORE DATA ANALYSIS.

Balanced

129S2				**BALB/c**				**C57BL/6**			
Dexmedetomidine		*Saline*		*Dexmedetomidine*		*Saline*		*Dexmedetomidine*		*Saline*	
Room K	Room L	Room K	Room L	Room K	Room L	Room K	Room L	Room K	Room L	Room K	Room L
8	8	8	8	6	6	6	6	6 (*-1*)	6	6 (*-1*)	6
16		*16*		*12*		*12*		*12 (-1)*		*12 (-1)*	
32				**24**				**24** (*-2*)			
80 *(-2)*											

TABLE 7. OVERVIEW OF SAMPLE SIZE OF THE GROUP OF ANIMALS THAT WAS UNBALANCED IN TERMS OF RESPONSE TYPE CLUSTER. "(-N)" INDICATES THE DECISION TO EXCLUDE N ANIMAL(S) IN THIS GROUP/SUBSET AFTER FINISHING THE EXPERIMENT AND DATA COLLECTION, BUT BEFORE DATA ANALYSIS.

Unbalanced

129S2				**BALB/c**				**C57BL/6**			
Dexmedetomidine		*Saline*		*Dexmedetomidine*		*Saline*		*Dexmedetomidine*		*Saline*	
Room K	Room L	Room K	Room L	Room K	Room L	Room K	Room L	Room K	Room L	Room K	Room L
6	7	6	7	6	7	6	7	6 (*-1*)	8	6 (*-1*)	8
13		*13*		*13*		*13*		*14 (-1)*		*14 (-1)*	
26				**26**				**28** (*-2*)			
80 *(-2)*											

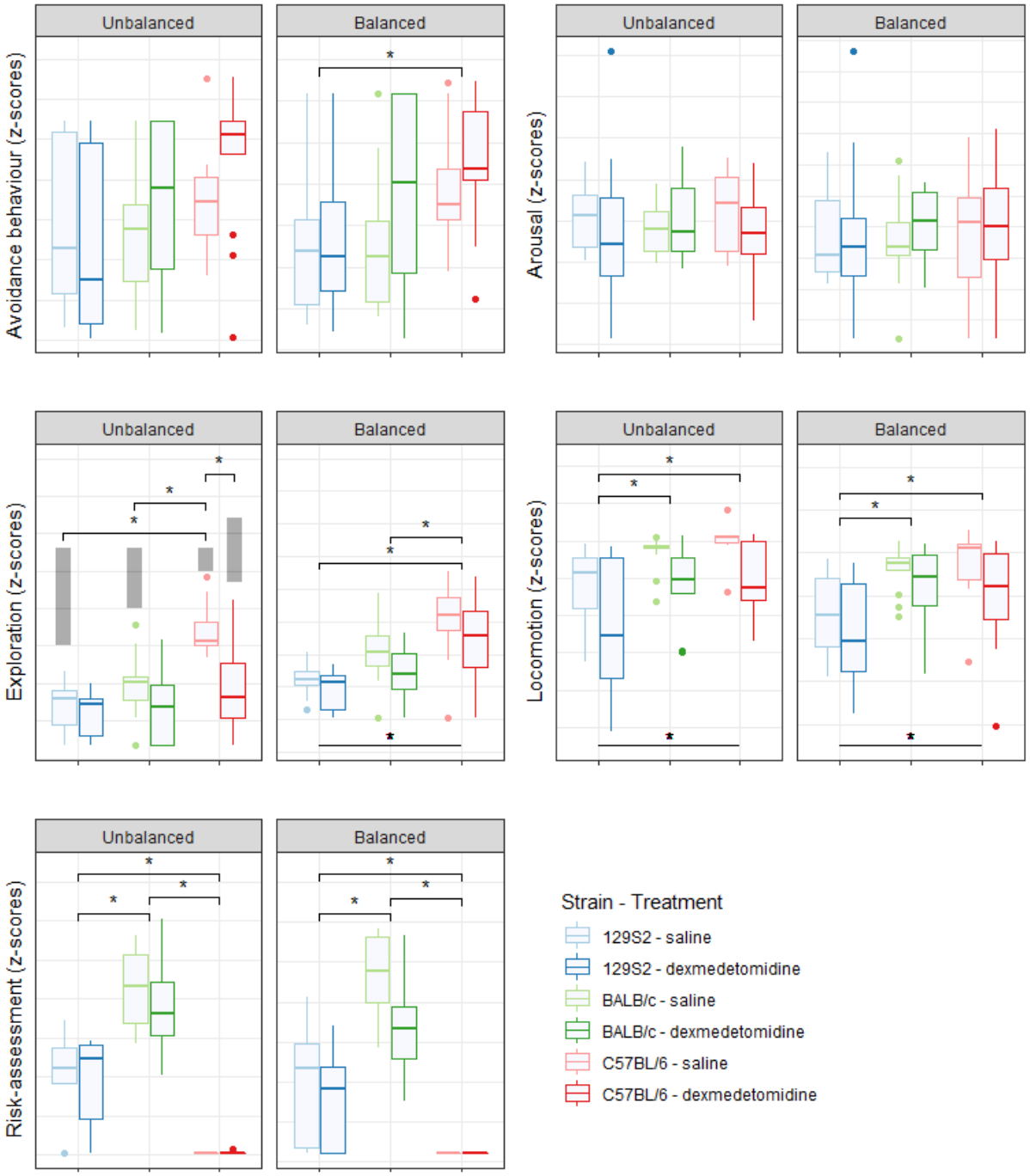


FIGURE 8. OVERVIEW OF TRANSFORMED Z-SCORES PER BEHAVIOURAL DIMENSION IN BOXPLOTS. UNBALANCED AND BALANCED BOXPLOTS ARE SEPARATED, SINCE ALL Z-SCORING WAS DONE SEPARATELY ON THE BALANCED AND UNBALANCED GROUP. DIFFERENT COLOURS INDICATE THE THREE DIFFERENT INBRED STRAINS, IN WHICH THE LIGHTER COLOURS INDICATE INTRAPERITONEALLY INJECTED SALINE AS TREATMENT, WHEREAS THE DARKER COLOURS INDICATE INTRAPERITONEALLY INJECTED DEXMEDETOMIDINE AS TREATMENT. COLOURED DOTS SIGNIFY OUTLIERS. "*" PRESENTS A SIGNIFICANT RESULT IN PLANNED COMPARISONS. BLACK SQUARE BRACKETS ON TOP OF BOXPLOTS INDICATE THAT THE SIGNIFICANT RESULT IS BETWEEN TWO STRAINS, REGARDLESS OF TREATMENT; SQUARE BRACKETS OVER TRANSLUCENT GREY BARS INDICATE THAT THE SIGNIFICANT RESULT IS BETWEEN TWO STRAINS WITHIN A SPECIFIC TREATMENT; BLACK LINES UNDERNEATH BOXPLOTS INDICATE THAT THE SIGNIFICANT RESULT IS IN TREATMENT REGARDLESS OF STRAIN. TRANSLUCENT GREY BARS ARE ONLY USED AS MEANS TO ELUCIDATE ON THE INDICATED SIGNIFICANT RESULTS: THEY ARE NOT LINKED TO THE QUALITATIVE DATA.

The data as provided in figure 8 is divided into group (balanced vs. unbalanced), strain (129S2 vs. BALB/c vs. C57BL/6) and treatment (dexmedetomidine vs. saline). Note that no comparison of the z-scores between the balanced and unbalanced groups was made, as the z-scores have been created per test group.

4.3.1 Avoidance behaviour

Considering the behavioural dimension of avoidance in the balanced group, experimenter, strain and treatment significantly predicted avoidance behaviour (respectively: χ^2 (1) = 6.66, p = 0. 0099; χ^2 (2) = 10.98, p = 0.0041; χ^2 (1) = 4.70, p = 0.0301). Pairwise comparisons revealed only a significant difference between the 129S2 strain and the C57BL/6 strain (|0.603| ± 0.229 (non-transformed), p = 0.0162), which showed that the C57BL/6 mice scored higher in their avoidance behaviour than the 129S2 mice.

The unbalanced group showed corresponding significant results in experimenter (χ^2 (1) = 8.01, p = 0.0047), but did not show significant results in strain nor treatment. Through planned comparison, the results on experimenter did not turn out significant. Table 8 gives an overview of all planned comparisons in avoidance behaviour, their p-value, their effect size and their effect size correlation. The unbalanced model was revealed to have different results than the balanced model.

Table 8. Planned comparison in the balanced and unbalanced groups in avoidance behaviour. "-" indicates a between-comparison between two groups. The presented estimate and standard error (SE) are described in non-transformed format. "*" indicates a significant result in the comparison.

Comparison	Estimate	SE	P-value	Cohen's d	Contrast r
					Balanced
129S2 - BALB/c	-0.242	0.229	0.7925	-0.391	0.192
129S2 - C57BL/6	-0.603	0.229	*0.0162	-1.104	0.483
BALB/c - C57BL/6	-0.362	0.232	0.2321	-0.704	0.332
Room K - Room L	0.417	0.188	0.0850	0.870	0.399
Saline - Dexmedetomidine	-0.311	0.141	0.1943	-0.672	0.319
					Unbalanced
Room K - Room L	0.478	0.183	0.0098	0.954	0.430

4.3.2 Arousal

In regards to arousal, no significant effects in the fixed factors experimenter, strain and treatment were found in the balanced group nor in the unbalanced group. Therefore, both the balanced and unbalanced group presented similar results.

4.3.3 Exploration

The balanced group showed significant results concerning the exploration dimension: strain (χ^2 (2) = 30.95, p < 0.0001) as well as treatment (χ^2 (1) = 12.41, p = 0.0004) significantly predicted exploration in the balanced group, although experimenter did not reveal any significant influence. The results in the planned comparison showed that treatment with dexmedetomidine caused a decrease in exploration (|0.257| ± 0.080 (non-transformed), p = 0.0061), whereas it was shown that the C57BL/6 mice showed more exploration than the 129S2 and BALB/c strains (respectively: |0.860| ± 0.148 (non-transformed), p < 0.0001; |0.540| ± 0.150 (non-transformed), p = 0. 0158). No significant difference was found between the other two strains.

On the other hand, a significant interaction in exploration between strain and treatment was found in the unbalanced group (χ^2 (2) = 9.36, p = 0.0093). Estimated marginal means revealed that saline-

treated C57BL/6 mice showed significantly higher levels of exploration compared to saline-treated 129S2 and BALB/c mice (respectively: |1.10| ± 0.158 (non-transformed), p < 0.0001; |0.809| ± 0.161 (non-transformed), p = 0.0007). Such significance was not found between the dexmedetomidine-treated strains. In addition to this, the unbalanced C57BL/6 mice presented a significant difference in treatment, in which saline-treatment showed lower exploration z-scores than the dexmedetomidine-treatment (|0.785| ± 0.161 (non-transformed), p = 0.0004).

All aforementioned results on exploration can be found in table 9, as well as effect size and effect size correlation from both models. Similarly to the avoidance behaviour dimension, the balanced and unbalanced group show different results regarding exploration.

TABLE 9. PLANNED COMPARISON IN THE BALANCED AND UNBALANCED GROUPS IN EXPLORATION. "-" INDICATES A BETWEEN-COMPARISON BETWEEN TWO GROUPS. THE PRESENTED ESTIMATE AND STANDARD ERROR (SE) ARE DESCRIBED IN NON-TRANSFORMED FORMAT. "*" INDICATES A SIGNIFICANT RESULT IN THE COMPARISON.

Comparison		Estimate	SE	P-value	Cohen's d	Contrast r
						Balanced
129S2 - BALB/c		-0.319	0.148	0.1250	-0.776	0.362
129S2 - C57BL/6		-0.860	0.148	*<0.0001	-1.869	0.683
BALB/c - C57BL/6		-0.540	0.150	*0.0158	-1.078	0.475
Saline - Dexmedetomidine		0.257	0.080	*0.0061	1.092	0.479
						Unbalanced
Saline	129S2 - BALB/c	-0.290	0.158	0.4262	-0.679	0.321
Saline	129S2 - C57BL/6	-1.099	0.158	*<0.0001	-2.286	0.752
Saline	BALB/c - C57BL/6	-0.809	0.161	*0.0007	-1.577	0.619
Dexmedetomidine	129S2 - BALB/c	-0.091	0.158	0.9998	-0.178	0.089
Dexmedetomidine	129S2 - C57BL/6	-0.401	0.158	0.1569	-0.860	0.395
Dexmedetomidine	BALB/c - C57BL/6	-0.310	0.161	0.4426	-0.670	0.318
Saline - Dexmedetomidine	129S2	0.087	0.155	0.9983	0.220	0.109
Saline - Dexmedetomidine	BALB/c	0.286	0.161	0.3501	0.676	0.320
Saline - Dexmedetomidine	C57BL/6	0.785	0.161	*0.0004	1.533	0.608

4.3.4 LOCOMOTION

Locomotion was found to show significant strain and treatment effects (respectively: χ^2 (2) = 16.21, p = 0.0003; χ^2 (1) = 10.73, p = 0.0011) in the balanced group. Similar results were revealed in the planned comparison: the 129S2 animals showed significantly lower z-scores in the locomotion dimension compared to the BALB/c and C57BL/6 animals (respectively: |0.534| ± 0.164, p = 0.0101; |0.564| ± 0.164, p = 0.0062), although the latter two did not show significant differences between each other. A significant negative influence of dexmedetomidine on locomotion was also found (|0.291| ± 0.092, p = 0.0117). Similar to the balanced group, significant differences in the unbalanced group's locomotor behaviour between strains and between treatments were found (respectively: χ^2 (2) = 24.80, p < 0.0001; χ^2 (1) = 20.03, p < 0.0001). Again, the 129S2 presented lower levels of locomotor behaviour compared to the BALB/c and C57BL/6 strains (respectively: |0.493| ± 0.135, p = 0.00035; |0.608| ± 0.135, p = 0.0003) and saline-treated animals scored higher in their locomotor activity than dexmedetomidine-treated animals (|0.479| ± 0.111, p = 0.0004). As can be seen in table 10, which shows all planned comparison results on locomotion, no differences were found between the balanced and unbalanced group concerning locomotion.

TABLE 10. PLANNED COMPARISON IN THE BALANCED AND UNBALANCED GROUPS IN LOCOMOTION. "-" INDICATES A BETWEEN-COMPARISON BETWEEN TWO GROUPS. "*" INDICATES A SIGNIFICANT RESULT IN THE COMPARISON.

Comparison	Estimate	SE	P-value	Cohen's d	Contrast r
					Balanced
129S2 - BALB/c	-0.534	0.164	*0.0101	-1.137	0.494
129S2 - C57BL/6	-0.564	0.164	*0.0062	-1.200	0.515
BALB/c - C57BL/6	-0.023	0.166	0.9996	-0.062	0.031
Saline - Dexmedetomidine	0.291	0.092	*0.0117	1.016	0.453
					Unbalanced
129S2 - BALB/c	-0.493	0.135	*0.0035	-1.275	0.538
129S2 - C57BL/6	-0.608	0.135	*0.0003	-1.571	0.618
BALB/c - C57BL/6	-0.114	0.137	0.8787	-0.291	0.144
Saline - Dexmedetomidine	0.479	0.111	*0.0004	1.388	0.570

4.3.5 RISK-ASSESSMENT

Strain and treatment significantly predicted risk-assessment in the balanced group (respectively: χ^2 (2) = 216.35, $p < 0.0001$; χ^2 (2) = 6.2507, $p = 0.0124$). Through estimated marginal means, it was found that all strains differed significantly from each other in this behavioural dimension, but no significant results were found again in treatment. BALB/c mice scored significantly higher on risk-assessment than the 129S2 and the C57BL/6 strains (respectively: $|1.122| \pm 0.133$, $p < 0.0001$; $|1.914| \pm 0.135$, $p < 0.0001$) and 129S2 mice scored significantly higher compared to the C57BL/6 mice ($|0.792| \pm 0.133$, $p < 0.0001$). Identical to this were the results found on risk-assessment in the unbalanced group: inter-strain difference was the only factor in the model to score significantly (χ^2 (2) = 263.51, $p < 0.0001$), the BALB/c mice again showed significantly highest z-scores on risk-assessment compared to the 129S2 and C57BL/6 mice (respectively: $|1.043| \pm 0.123$, $p < 0.0001$; $|1.985| \pm 0.126$, $p < 0.0001$) and the 129S2 animals presented significantly higher levels of risk-assessment compared to the C57BL/6 animals ($|0.942| \pm 0.124$, $p < 0.0001$). Table 11 shows all planned comparison results on risk-assessment, in which the balanced and unbalanced group are complementary to each other.

TABLE 11. PLANNED COMPARISON IN THE BALANCED AND UNBALANCED GROUPS IN RISK-ASSESSMENT. "-" INDICATES A BETWEEN-COMPARISON BETWEEN TWO GROUPS. "*" INDICATES A SIGNIFICANT RESULT IN THE COMPARISON.

Comparison	Estimate	SE	P-value	Cohen's d	Contrast r
					Balanced
129S2 - BALB/c	-1.122	0.133	*<0.0001	-2.93	0.826
129S2 - C57BL/6	0.792	0.133	*<0.0001	2.07	0.719
BALB/c - C57BL/6	1.914	0.135	*<0.0001	4.93	0.927
Saline - Dexmedetomidine	0.264	0.109	0.0785	0.76	0.361
					Unbalanced
129S2 - BALB/c	-1.043	0.123	*<0.0001	-2.94	0.827
129S2 - C57BL/6	0.942	0.124	*<0.0001	2.65	0.798
BALB/c - C57BL/6	1.985	0.126	*<0.0001	5.47	0.939

4.3.6 F-TESTS ON VARIANCE IN RESIDUALS

F-tests to compare two variances were also executed between the residuals of the balanced and unbalanced groups per behavioural dimension, but none of these five dimensions demonstrated significant results in variance of the residuals when comparing the balanced group to the unbalanced group.

5. DISCUSSION

To summarise, the present study found that taking inter-individual differences into account when creating test groups led to different results in the behavioural dimensions of exploration and avoidance behaviour compared to when these differences were not taken into consideration. No similar differences in results were found in the behavioural dimensions of locomotion, arousal and risk-assessment. Additionally it was found that even when comparable significant factors were found within the same behavioural dimension, effect size (in the form of Cohen's d and the contrast r) always yielded higher results in the groups unbalanced in response type compared to the groups balanced in response type. The following chapter will first discuss the results regarding the research question and will then address other results and obstacles found during the present study.

5.1 THE EFFECTS OF CONTROLLING INTER-INDIVIDUAL DIFFERENCES ON RELIABILITY OF RESEARCH

First, it can be stated that the use of standardisation in behavioural research is a necessity and that confounding factors need to be taken into account when executing behavioural studies (75,76). Consequently, analysing behavioural patterns and taking said patterns into consideration when dividing the animals into groups, could be used as a means to increase standardisation. The question was whether or not identifying response types in individuals would contribute to standardisation in behavioural testing. The results provided in the present study indicate that the method as it was implemented can influence the results and could improve reliability and reproducibility in future behavioural research: the results provided by the balanced group in the avoidance and exploration dimensions differed from the results provided by the unbalanced group, meaning that they showed different factors to significantly affect the results. Such dissimilarity was, however, not found when comparing the arousal, risk-assessment and locomotion between both groups.

Provided that arousal and risk-assessment were also not taken into account in the cluster analysis, it was conceivable that these behavioural dimensions did not show differences between the balanced and unbalanced group. The fact that locomotion showed the same results in both the balanced and unbalanced group might be explained by the assumption that the results were influenced by the factors to such an extent that the inter-individual variation did not affect the degree of significance in prediction of the factors. Nevertheless, when comparing similar significant results between the balanced and unbalanced groups in both locomotion and risk-assessment, Cohen's d and the contrast r always returned higher outcomes in the unbalanced group than was found in the balanced group. It is possible that this is caused by an over-estimation of the results in the unbalanced group, leading to higher effect sizes. Along the same lines, this could also explain how, in the exploration dimension, the two-way interaction of treatment and strain yielded significant results in the unbalanced group, but not in the balanced group. Avoidance behaviour, on the other hand, demonstrated more significant results in the balanced group, possibly indicating that taking individual variation into account decreased the variance in the results. However, validation for this assumption was not found by the F-test on variance in residuals of both the balanced and unbalanced model.

It can be noted that no analysis on response type cluster was done in phase 2 of the present study. Response type clusters should in the basis be used as a means to increase standardisation within a study. As stated before, it has been demonstrated numerous times that behavioural inter-individual differences are multifactorial and approaches to control said differences have yet to be developed. The method presented in the present study provides a flexible way of controlling results that are influenced by the inter-individual differences, meaning that this method could be used on another sample of individuals, provide different response types clusters and still improve reliability and repeatability

of the study. Consequently, analysing the results in phase 2 by using cluster as an independent variable does not only carry little weight, it also degrades the method and its usage. In other words, results obtained from analysis on response type cluster may not be repeatable and are thus not adequately reliable. On the other hand, it might be interesting to utilise inferential statistics on the results of phase 1 to investigate inter-cluster differences, since this could increase understanding of the usage of longitudinal data clustering to identify inter-individual differences.

Overall, the in phase 2 acquired results of the balanced group show that 1) 129S2 mice showed significantly lower avoidance behaviour compared to the C57BL/6 group; 2) C57BL/6 mice showed significantly higher z-scores in exploration compared to both other strains; 3) 129S2 mice showed significantly lower levels of locomotion compared to both other strain and 4) all strains differed from each other in risk-assessment, with the BALB/c strain showing the highest scored, followed by the 129S2 strain and with the C57BL/6 strain showing the lowest scores. First of all, the C57BL/6 has previously demonstrated to show more avoidance-like behaviour when socially isolated (77,78). However, no comparison between C57BL/6 mice and 129S2 has been done before, so there is no evidence provided to explain this difference. Secondly, the C57BL/6 and BALB/c mice have been compared in research before, which supported the result that C57BL/6 animals show high exploratory activity in comparison to the BALB/c animals (79). Nevertheless, the C57BL/6 and BALB/c strain have yet to be statistically compared to the 129S2 strain regarding exploration. Studies have shown, on the other hand, that 129S2 mice show less locomotor activity than C57BL/6 mice (80) and that intersession habituation leads to a decrease of locomotor behaviour in 129S2 and C57BL/6 mice, whereas it leads to an increase in BALB/c mice (81). No significant difference between C57BL/6 and BALB/c mice regarding locomotion was found in the present study, although such a difference has been demonstrated in previous research (38,47,48,79). Lastly, the results in risk-assessment are completely in line with previous research, which shows that both the BALB/c and 129S2 strains display more risk-assessments compared to the C57BL/6 strain (79,80,82,83). BALB/c mice and 129S2 mice have not previously been analysed on inter-strain differences. Therefore, no evidence was available to support the risk-assessment difference between these two strains.

To conclude, since avoidance behaviour and exploration showed different results between the balanced and unbalanced group, the reliability of the results in the present study is deemed to be improved by taking inter-individual variation in response type into account.

5.2 OTHER RELEVANT MATTERS OF INTEREST

Over the course of this experiment, several obstacles were encountered and supplementary results were found. Said obstacles are presented to explain several results, as well as to create more coherent future directions. To create a clear overview of the segments of this experiment that were suboptimal and additional results that have been obtained, other than those answering the research question, discussion of these components will be done in four sections. These are sections on the experimenter effects, the reliability of results regarding treatment during the pilot and phase 2, the corticosterone samples and their analysis and the clusters and establishment thereof. It should be noted that, since the previous paragraph demonstrated how the usage of clusters seemingly improved standardisation in the present experiment, only the balanced group will henceforth be taken into account when discussing results.

5.2.1 EXPERIMENTER EFFECTS

The experimenter factor, as described in the linear mixed models, did not only account for the possible difference in behavioural testing between the two experimenters, regardless of preceding training for unison; it also accounted for the possible environmental differences between the two rooms. Although

standardisation was considered paramount during the present study, the amount of animals did not allow for the animals to be tested by one experimenter nor housed in one room. As a result, any diversity in temperature of air humidity between the two rooms is also included in the experimenter factor. As mentioned before, the average temperature and air humidity in room K were 21.7 °C (± 0.234) and 53.5% (± 2.44), whereas the average temperature and air humidity in room L were 21.9 °C (± 0.403) and 54.8% (± 2.56). These means and standard deviations show that the rooms did not differ in living conditions on average. Appendix II shows boxplots of the overall temperature and humidity per room and line graphs of the temperature and humidity over time per room.

Furthermore, a noticeable strain-difference in weight-gain was found over time: whilst the 129S2 and the BALB/c strains showed ordinary body condition scores (scores between 3 and 4) and demonstrated consistent weight-gain, some C57BL/6 mice displayed excessive weight gain with body condition scores later reaching between 4 and 5, reaching obese levels over time (58–60). Individuals within the C57BL/6 strain exclusively showed these conditions. Other studies have shown that long-term corticosterone exposure decreases insulin sensitivity in this specific C57BL/6 strain (84) and that weight gain may be a result of social isolation in mice (85,86). As a result, it may be possible that social isolation caused this specific body condition in the C57BL/6 strain and/or that the C57BL/6 were exposed to chronically elevated corticosterone levels due to social isolation. However, the latter is unlikely, due to various studies having investigated the influence of social isolation on corticosterone levels in mice and deeming it non-significant (55,87,88), in addition to the present study's results demonstrating significantly lower corticosterone levels in C57BL/6 mice compared to the other two inbred strains.

Additionally, the experimenters observed that several BALB/c mice showed excessive aggressive behaviour towards the experimenters during behavioural testing, handling and weighing. Past research has shown that previously isolated BALB/c mice show more aggressive behaviour towards other mice when re-socialised (89). It has been theorised that this aggression may have derived from territorial aggression (90), which has shown to be influenced by the serotonergic system (91–94). More specifically, serotonin was demonstrated to be negatively affect levels of aggressivity (91–94), whilst research has also revealed how BALB/c mice show reduced levels of brain serotonin (95). To conclude, the social isolation may have led to a decrease in brain serotonin (96–98), which led to the display of human-directed aggression in the BALB/c strain.

It is worth mentioning that, although some studies have shown influence of social isolation (77,78), such effects of social isolation have been studied in the mHB, but it was demonstrated that social isolation did not seem to affect the results (55). Hence, the influence of social isolation on the results of the present study was deemed insignificant.

5.2.2 Reliability of results in treatment effect

The results have shown that the treatment factor significantly influenced z-scores in exploration and locomotion in the group balanced on response type. These results imply that dexmedetomidine caused a decrease in exploration and locomotion. No effects of the treatment factor were found in risk-assessment and avoidance behaviour. This was unexpected, since the use of an anxiolytic dose of dexmedetomidine was thought to cause an increase in exploratory behaviour and not influence locomotion. It was also anticipated that dexmedetomidine would increase risk-assessment and avoidance behaviour. The obtained results suggest the dexmedetomidine injection to be anxiogenic, in contrast to the desired anxiolytic effect. An explanation for this could be that the expectations on treatment in mice with this drug were not accurate. Previous studies indicate that, rather than locomotor activity remaining unaffected, dexmedetomidine causes a decrease in locomotor activity (99,100). Nevertheless, this does not readily explain the absence of effect on the other behavioural dimensions, since previous studies also indicate the anxiolytic effects of dexmedetomidine (69,99,101,102). Note that

the apparent anxiogenic effects of dexmedetomidine by using this dose did not influence the reliability of the results regarding the main hypothesis, since the main research question was to investigate improving reliability when using a different, more standardised test set-up in a medicinal testing environment. Since the same dose has been used in both test set-ups, the reliability of the results regarding the main hypothesis remains unaffected.

Another possibility for the results diverging from the predictions, is that the pilot did not provide the correct dose. As only five animals were used per strain to test three different doses, the pilot data may not have been reliable enough to support the choice of the dose with statistical analysis. Additionally, the use of the more selective adrenergic receptor agonist dexmedetomidine over medetomidine as an anxiolytic in mice has yet to be studied extensively in the behavioural field of research (70). All in all, this provided little evidence to determine the required dose. Furthermore, the difference of age between the animals in the pilot study and the animals in phase 2 was over ten weeks, with the mice in the pilot study being between nine and ten weeks old and the mice that were injected in phase 2 being over 20 weeks old. It is possible that the effective dose of dexmedetomidine in mice varies per age (103).

What also should be considered is that the decision was made to use the same dose in all three strains. This approach may have been inaccurate, since different mouse strains have demonstrated differences in α_2-adrenergic receptor binding (104,105). As mentioned before, inter-strain differences have already been established regarding serotonin levels. Dexmedetomidine has been demonstrated to influence central serotonin levels to some extent in rats and mice (106–110). It might thus be more accurate to establish the effective dose per strain, instead of using the same dose in all three strains.

Finally, it was noted by the experimenters that eight out of 26 C57BL/6 mice (30.8%) that were treated with dexmedetomidine in phase 2 seemed to display hypertonia in the limbs during testing. Similar results have not been documented before, with some studies even describing the prevention of opiate-induced muscle rigidity in rats by dexmedetomidine (111,112). However, a spontaneously mutated spasticity gene has previously been reported in C57BL/6 mice (113) and the aforementioned excessive weight gain in the C57BL/6 animals could also have affected the results, since dexmedetomidine is a highly lipophilic drug (101) and the predicted steady-state concentrations were demonstrated to be significantly influenced by obesity in human patients (102). These observations only add to the statement that creating strain-dependent doses may improve the reliability of future research in behavioural studies using dexmedetomidine as an anxiolytic.

5.2.3 Corticosterone measurements

Although corticosterone was not taken into account in creating the response types, corticosterone levels were analysed on inter-strain and inter-cluster differences separately. The acquired results from these analyses showed that testing caused an increase in corticosterone compared to pre-testing samples in all three strains, as was anticipated. A stressful situation, such as the five times five minutes during which a mouse was placed in a novel environment through handling, would cause the hippocampal-pituitary-adrenal axis to activate and corticosteroids to be released (114). As the release of corticosterone involves hormonal processes (115), it is to be expected that the first and third blood sampling moment affected the corticosterone levels less, since the stressful situation during the first and third blood sampling moment was created by moving the mouse in its home cage to another room, after which blood was taken. However, the significant difference in corticosterone levels between the first and third sampling moment in 129S2 mice cannot readily be explained. Not much research has been conducted on 129S2 substrains; consequently, no other studies show a similar set-up. The results as presented imply the probability that 129S2 mice have not been able to return to their pre-testing levels after one week, but such assumptions could not be made from previous research (23,24).

The results showed that BALB/c animals had significantly higher levels of corticosterone than C57BL/6 animals on each sampling moment, an observation that is consistent with other findings (116,117). It has been shown that BALB/c mice show lower levels of mineralo- and glucocorticoid receptor mRNA in the limbic brain compared to C57BL/6 mice, whilst also presenting a more rapid and higher corticosterone release compared to the C57BL/6 strain when placed in a novel environment (116). As a result of the lower levels of mineralo- and glucocorticoid receptor mRNA in the limbic brain, the negative feedback system that is present in the hypothalamus-pituitary-adrenal axis would have a smaller influence in the BALB/c strain than in the C57BL/6 strain, allowing for a higher and more rapid release of corticosterone when faced with stressful situations (118,119). Nonetheless, other studies have reported no such differences in basal corticosterone levels between BALB/c and C57BL/6 mice (117,120,121). Consequently, it may be possible that the supposed baseline levels had already been affected by moving the cage (122), which could explain the higher baseline levels found in BALB/c compared to C57BL/6.

Next, when comparing the response type clusters, cluster A demonstrated results similar to strain 129S2: all sampling moments differed significantly from each other. Response type B showed only the peak of the second sampling moment to be significant. This is completely in line with the previously mentioned inter-strain differences in corticosterone levels per sampling moment, as the majority of 129S2 mice were assigned to cluster A, thereby influencing the results with inter-strain differences. A model including strain and cluster had also been analysed, but was deemed to overfit the data. Nonetheless, it did only showed a significant strain-sampling moment interaction, foregoing the implication of a cluster-sampling interaction. It can thus be concluded that the cluster-sampling moment interaction was only found when excluding strain, due to the distribution and resulting influence of strain on the response type cluster.

It should be mentioned that the reliability of the acquired corticosterone levels can be considered dubious. The low and high control samples that were used to create the calibration curves, as supplied by the manufacturer, turned out higher than the provided control range in almost all centrifuge-run batches. This means that the provided results may be higher than the actual corticosterone levels or may even be non-reliable at all, as the manufacturer's criteria for reliable results were not met.

5.2.4 Cluster analysis in previous studies and in future directions

The cluster analysis combined with the three selection criteria yielded two clusters as the most optimal distribution of the individuals. However, as mentioned before, two out of the five initially analysed behavioural dimensions – arousal and risk-assessment – as well as the corticosterone levels were left out of this analysis. Results in phase 2 were on par with this, since no different results between the balanced and unbalanced groups have been found.

Prior to the risk-assessment dimension being left out of the cluster analysis, its residuals had been obtained by using a different type of model compared to the other four behavioural dimensions. The risk-assessment behaviour showed a zero-inflation, resulting in a linear mixed model that could not be normally distributed, despite transformation attempts. This zero-inflation was caused by a strain-difference overall and per trajectory: whereas the other two strains showed higher amounts of risk-assessments during the first and second trial, C57BL/6 mice showed barely any risk-assessments during any trial. Moreover, the other two strains drastically decreased their amounts of risk-assessments after the first two trials. Other papers have also noted this difference between the C57BL/6 strain and the BALB/c and 129S2 strains (79,80,82,83) and the effect of number of trials on the risk-assessment paradigm has been demonstrated by other studies (23,24). Because the risk-assessment dimension was computed out of a single behaviour vector, which included two variables (latency of first risk-assessment and total amount of risk-assessments), zero counted values from all three strains were able to

have a large effect on this dimension. By including zero-inflation in the model and utilising a gamma-distribution, the zero-inflation was accounted for, resulting in a normally distributed linear mixed model (123).

In addition to the C57BL/6 strain being the cause of zero-inflated data in the risk-assessment dimension, the distribution of the strains amongst the two clusters is noteworthy: as shown in table 1, 88.3% of the 129S2 animals were allocated in cluster A, whereas 84.7% of the BALB/c animals were placed in cluster B. Both strains are also shown to dominate these respective clusters. The total distribution of mice among the two response type clusters seems relatively equal. Similar results have been found in Exp. 1 (26), in which the BALB/c strain was allocated mostly to a certain response type and was dominating this cluster, with 62.5% of the BALB/c mice being placed in one specific response type cluster and the cluster consisting for 75% of BALB/c mice. All of these results are consistent with the fact that, although they were accounted for in the linear mixed models, inter-strain differences make up for the majority of the behavioural differences found between groups of different individuals (8–12,22,50,74,80,82,83,123). However, it shows that not all differences can be accounted for by strain and, as a result, the use of cluster analysis may be able to improve reliability of behavioural research.

Moreover, the preceding experiment's response type clusters (26) were compared to the present study's response type clusters and it was found that, although the finalised amount of clusters was three in Exp. 1, said three response types show similarities to the response types shown in the present study. More specifically, response type B in the present study and response type C "Habituation" in the preceding study showed several analogies: both showed residual increases in exploration and residual decreases in avoidance behaviour over time compared to the other clusters. Arousal was also relatively in line between the response types from the two studies, staying somewhat unchanged over time. On the other hand, locomotion had not previously presented the increase as was shown in the present study and whilst risk-assessment seemed to look unaffected by the number of trials, it had presented a decrease in response type C in the foregoing experiment. Similarities between response type A in the study and response type A "Non-responding" and/or response type B "Sensitisation" were not as outspoken. (26)

Regarding future research, a recent study has demonstrated how copy number variation in small nucleolar RNAs regulates personality behaviour in, inter alia, C57BL/6 mice (17). It showed how certain RNA clusters displayed copy number variation and affected the expression of specific genes; a strong correlation was found between said RNA clusters and scores in behavioural tests. Since the so-called snoRNAs might be (partly) responsible for inter-individual variation in behaviour (17), future studies investigating k-means clustered data could consider combining the data with data on the individuals in copy number variation in these small nucleolar RNAs. This might increase understanding in inter-individual differences in emotional behaviour and how this variation has come into existence, even in genetically similar laboratory rodent strains.

6. CONCLUSION AND FUTURE DIRECTIONS

In summary, the present study has shown that inter-individual differences may be standardised through the use of k-means clustering for longitudinal data on response type. The provided results focused on the variation in behavioural responses between strains in order to create the clusters; afterwards, pairs were created based on these results. Avoidance behaviour and exploration models showed different results with significant p-values when comparing balanced pairs to pairs that were not created based on response type cluster. However, the behavioural dimensions of risk-assessment and arousal were not taken into account when creating the response type clusters and also did not reveal a difference between cluster-standardised or non-cluster-standardised individuals in phase 2 of the present study. Whenever significant p-values had been yielded in the same factors in both the balanced and unbalanced group, effect sizes were found to be higher in factors in the unbalanced group.

Corticosterone was left out of the analysis as well and was determined to be dubious in its reliability, due to the control samples deviating from the criteria provided by the manufacturer. An independent analysis on inter-strain differences in corticosterone levels revealed that the BALB/c strain showed higher levels of corticosterone than the C57BL/6 strain and that the 129S2 strain was implied not to be able to recover as quickly from a stressful situation regarding corticosterone levels.

Furthermore, the effects of dexmedetomidine in mice and between different strains should be investigated. Additionally, determining strain-dependent doses of dexmedetomidine may improve the quality of behavioural research when using this drug. More research should also be conducted on the strain-dependent behavioural differences, since not all findings in the present study could be validated. For example, few studies have been published that conduct inter-strain comparisons between 129S2 mice and BALB/c and/or C57BL/6 mice and the present study was thus not able to validate all findings on inter-strain differences.

To conclude, reliability of behavioural results in the present study has been improved by taking inter-individual variation of emotional reactivity into account in the composition of experimental groups of mice. When using this method, it is not necessary to add a group that has not been balanced in regards to response type cluster, since the present study has already provided results with the use of p-values and effect sizes that indicate that balancing groups on response type is more reliable.

After thorough analyses of the results produced by the present study, further future directions may contain necessary validation of the results by replication or extending similar research to outbred stocks, female mice or other laboratory rodents. It can also be considered to investigate correlation between k-means clustered data based on intersession habituation vs. copy number variation in small nucleolar RNAs. Validation of the presented method can be conducted readily on other standardised groups of mice, due to the flexibility of the method. Hence, the use of this method may be able to contribute to a reduction of a required sample size in the future.

REFERENCE LIST

1. Georgejason. White mouse for scientific research stock photo · baby mouse on black background stock photo [Internet]. [cited 04-07-0219]. Available from: https://media.istockphoto.com/photos/white-mouse-for-scientific-research-picture-id526539546?k=6&m=526539546&s=612x612&w=0&h=E1qn5a1SKU9PvOQr5v5nUqyhPtecK5u7Rr5omogvcgc=

2. Kilkenny C., Browne W.J., Cuthill I.C., Emerson M., Altman D.G. Animal research: Reporting in vivo experiments: The ARRIVE guidelines. British Journal of Pharmacology. 2010 Aug 1;160(7):1577–9.

3. Pankevich D.E., Wizemann T.M., Altevogt B.M. Improving the Utility and Translation of Animal Models for Nervous System Disorders: Workshop Summary. National Academies Press; 2013.

4. Omary M.B., Cohen D.E., El-Omar E.M., Jalan R., Low M.J., Nathanson M.H., et al. Not All Mice Are the Same: Standardization of Animal Research Data Presentation. Cellular and Molecular Gastroenterology and Hepatology. 2016 Jul;2(4):391–3.

5. Bourin M. Animal models for screening anxiolytic-like drugs: a perspective. Dialogues in clinical neuroscience. 2015 Sep;17(3):295–303.

6. Ennaceur A., Chazot P.L. Preclinical animal anxiety research – flaws and prejudices. Pharmacology Research and Perspectives. 2016 Apr 1;4(2):e00223.

7. Ohl F. Testing for anxiety. Clinical Neuroscience Research. 2003 Dec 1;3(4–5):233–8.

8. Hufeldt M.R., Nielsen D.S., Vogensen F.K., Midtvedt T., Hansen A.K. Variation in the gut microbiota of laboratory mice is related to both genetic and environmental factors. Comparative medicine. 2010;60(5):336–47.

9. Loos M., Koopmans B., Aarts E., Maroteaux G., van der Sluis S., Verhage M., et al. Within-strain variation in behavior differs consistently between common inbred strains of mice. Mammalian genome. 2015;26(7–8):348–54.

10. Jensen V.S., Porsgaard T., Lykkesfeldt J., Hvid H. Rodent model choice has major impact on variability of standard preclinical readouts associated with diabetes and obesity research. American journal of translational research. 2016;8(8):3574.

11. Lewejohann L., Zipser B., Sachser N. “Personality” in laboratory mice used for biomedical research: A way of understanding variability? Developmental Psychobiology. 2011 Sep;53(6):624–30.

12. Koolhaas J.M., de Boer S.F., Coppens C.M., Buwalda B. Neuroendocrinology of coping styles: Towards understanding the biology of individual variation. Frontiers in Neuroendocrinology. 2010 Jul 1;31(3):307–21.

13. Button K.S., Ioannidis J.P.A., Mokrysz C., Nosek B.A., Flint J., Robinson E.S.J., et al. Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience. 2013 May 10;14(5):365–76.

14. Garner J.P. Stereotypies and other abnormal repetitive behaviors: Potential impact on validity, reliability, and replicability of scientific outcomes. ILAR Journal. 2005 Jan 1;46(2):106–17.

15. Freund J., Brandmaier A.M., Lewejohann L., Kirste I., Kritzler M., Krüger A., et al. Emergence of individuality in genetically identical mice. Science. 2013;340(6133):756–9.

16. Liu J., Wei W., Kuang H., Tsien J.Z., Zhao F. Heart rate and heart rate variability assessment

identifies individual differences in fear response magnitudes to earthquake, free fall, and air puff in mice. Lu L, editor. PLoS ONE. 2014 Mar 25;9(3):e93270.

17. Keshavarz M., Krebs-Wheaton R., Refki P., Guenther A. Copy number variation in small nucleolar RNAs regulates personality behavior. 2019;

18. Burokas A., Arboleya S., Moloney R.D., Peterson V.L., Murphy K., Clarke G., et al. Targeting the Microbiota-Gut-Brain Axis: Prebiotics Have Anxiolytic and Antidepressant-like Effects and Reverse the Impact of Chronic Stress in Mice. Biological Psychiatry. 2017 Oct 1;82(7):472–87.

19. Sandhu K. V., Sherwin E., Schellekens H., Stanton C., Dinan T.G., Cryan J.F. Feeding the microbiota-gut-brain axis: diet, microbiome, and neuropsychiatry. Translational Research. 2017 Jan 1;179:223–44.

20. Galatzer-Levy I.R., Bonanno G.A., Bush D.E.A., LeDoux J.E. Heterogeneity in threat extinction learning: substantive and methodological considerations for identifying individual difference in response to stress. Frontiers in Behavioral Neuroscience. 2013 May 29;7:55.

21. Russell W.M.S., Burch R.L., Hume C.W. The principles of humane experimental technique. Vol. 238. Methuen London; 1959.

22. van der Goot M.H., van Lith H.A., Ohl F. Identification of variability in response types in mice. In Abstracts Behavior 2015 – 34th Ethological Conference; 2015.

23. Boleij H., Salomons A.R., van Sprundel M., Arndt S.S., Ohl F. Not all mice are equal: welfare implications of behavioural habituation profiles in four 129 mouse substrains. PloS one. 2012;7(8):e42544.

24. Salomons A.R., Van Luijk J., Reinders N.R., Kirchhoff S., Arndt S.S., Ohl F. Identifying emotional adaptation: behavioural habituation to novelty and immediate early gene expression in two inbred mouse strains. Genes, Brain and Behavior. 2010;9(1):1–10.

25. Genolini C., Alacoque X., Sentenac M., Arnaud C. kml and kml3d: R packages to cluster longitudinal data. Journal of Statistical Software. 2015;65(4):1–34.

26. van der Goot M.H., Arndt S.S., van Lith H.A. Increasing animal welfare and reliability of results from preclinical trials and animal studies - Zooming in on variation in adaptive response patterns within and between two mouse inbred strains. In 2016.

27. Ohl F., Holsboer F., Landgraf R. The modified hole board as a differential screen for behavior in rodents. Behavior Research Methods, Instruments, and Computers. 2001 Aug;33(3):392–7.

28. Labots M., van Lith H.A., Ohl F., Arndt S.S. The modified hole board--measuring behavior, cognition and social interaction in mice and rats. Journal of visualized experiments : JoVE. 2015 Apr 8;(98):52529.

29. Lister R.G. Ethologically-based animal models of anxiety disorders. Pharmacology and Therapeutics. 1990 Jan 1;46(3):321–40.

30. Koide T., Moriwaki K., Ikeda K., Niki H., Shiroishi T. Multi-phenotype behavioral characterization of inbred strains derived from wild stocks of Mus musculus. Mammalian Genome. 2000;11(8):664–70.

31. Montkowski A., Poettig M., Mederer A., Holsboer F. Behavioural performance in three substrains of mouse strain 129. Brain Research. 1997 Jul 11;762(1–2):12–8.

32. Labots M., Laarakker M.C., Schetters D., Arndt S.S., van Lith H.A. An improved procedure for integrated behavioral z-scoring illustrated with modified Hole Board behavior of male inbred

laboratory mice. Journal of Neuroscience Methods. 2018 Jan 1;293:375–88.

33. Guilloux J.P., Seney M., Edgar N., Sibille E. Integrated behavioral z-scoring increases the sensitivity and reliability of behavioral phenotyping in mice: Relevance to emotionality and sex. Journal of Neuroscience Methods. 2011 Apr 15;197(1):21–31.

34. Laarakker M.C., Ohl F., van Lith H.A. Chromosomal Assignment of Quantitative Trait Loci Influencing Modified Hole Board Behavior in Laboratory Mice using Consomic Strains, with Special Reference to Anxiety-related Behavior and Mouse Chromosome 19. Behavior Genetics. 2008 Mar 4;38(2):159–84.

35. Labots M. Chapter 6: Consomic mouse strain selection based on effect size measurement, statistical significance testing and integrated behavioral z-scoring: focus on anxietyrelated behavior and locomotion. In: Cracking anxiety in the mouse: a quantitative (epi)genetic approach. Utrecht University; 2017. p. 197–234.

36. Crawley J.N. Exploratory behavior models of anxiety in mice. Neuroscience and Biobehavioral Reviews. 1985 Mar 1;9(1):37–44.

37. Rogers D.C., Jones D.N.C., Nelson P.R., Jones C.M., Quilter C.A., Robinson T.L., et al. Use of SHIRPA and discriminant analysis to characterise marked differences in the behavioural phenotype of six inbred mouse strains. Behavioural Brain Research. 1999 Nov 15;105(2):207–17.

38. Ohl F., Sillaber I., Binder E., Keck M.E., Holsboer F. Differential analysis of behavior and diazepam-induced alterations in C57BL/6N and BALB/c mice using the modified hole board test. Journal of Psychiatric Research. 2001 May 1;35(3):147–54.

39. Festing M.F.W. Evidence should trump intuition by peferring inbred strains to outbred stocks in preclinical research. ILAR Journal. 2014 Dec 20;55(3):399–404.

40. Potter M. The BALB/c Mouse : Genetics and Immunology [Internet]. Springer Berlin Heidelberg; 1985. 253 p.

41. Peeler D.F., Nowakowski R.S. Genetic factors and the measurement of exploratory activity. Behavioral and Neural Biology. 1987 Jul 1;48(1):90–103.

42. Oliverio A., Eleftheriou B.E., Bailey D.W. A gene influencing active avoidance performance in mice. Physiology and Behavior. 1973 Oct 1;11(4):497–501.

43. Griebel G., Belzung C., Misslin R., Vogel E. The free-exploratory paradigm: an effective method for measuring neophobic behaviour in mice and testing potential neophobia-reducing drugs. Behavioural Pharmacology. 1993;4(6):637–44.

44. Merriam Webster. Emotion [Internet].

45. Simpson E.M., Linder C.C., Sargent E.E., Davisson M.T., Mobraaten L.E., Sharp J.J. Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. Nature Genetics. 1997;16(1):19–27.

46. Bryant C.D. The blessings and curses of C57BL/6 substrains in mouse genetic studies. Annals of the New York Academy of Sciences. 2011 Dec;1245(1):31–3.

47. Griebel G., Belzung C., Perrault G., Sanger D.J. Differences in anxiety-related behaviors and in sensitivity to diazepam in inbred and outbred strains of mice. Psychopharmacology. 2000 Feb 4;148(2):164–70.

48. Trullas R., Skolnick P. Differences in fear motivated behaviors among inbred mouse strains.

Psychopharmacology. 1993 Jun;111(3):323–31.

49. Kilkenny C., Browne W.J., Cuthill I.C., Emerson M., Altman D.G. Improving bioscience research reporting: The arrive guidelines for reporting animal research. Animals. 2013;4(1):35–44.

50. Morton D.B., Hau J. Welfare assessment and humane endpoints. In: Handbook of laboratory animal science. CRC Press Boca Raton; 2002. p. 457–86.

51. Palanza P., Gioiosa L., Parmigiani S. Social stress in mice: Gender differences and effects of estrous cycle and social dominance. Physiology & Behavior. 2001 Jun 1;73(3):411–20.

52. Palanza P., Parmigiani S. How does sex matter? Behavior, stress and animal models of neurobehavioral disorders. Neuroscience and Biobehavioral Reviews. 2017 May;76(Pt A):134–43.

53. Kästner N., Richter S.H., Gamer M., Kaiser S., Sachser N. What a difference a day makes—female behaviour is less predictable near ovulation. Royal Society Open Science. 2017 Apr 12;4(4):160998.

54. Fernández-Guasti A., Picazo O. Anxiolytic actions of diazepam, but not of buspirone, are influenced by gender and the endocrine stage. Behavioural Brain Research. 1997 Nov 1;88(2):213–8.

55. Arndt S.S., Laarakker M.C., van Lith H.A., van der Staay F.J., Gieling E., Salomons A.R., et al. Individual housing of mice — Impact on behaviour and stress responses. Physiology & Behavior. 2009 Jun 22;97(3–4):385–93.

56. Deacon R.M.J. Housing, husbandry and handling of rodents for behavioral experiments. Nature Protocols. 2006 Aug;1(2):936–46.

57. Rowan A.N. Refinement of animal research technique and validity of research data. Fundamental and applied toxicology : official journal of the Society of Toxicology. 1990 Jul;15(1):25–32.

58. Ullman-Culleré M.H., Foltz C.J. Body condition scoring: a rapid and accurate method for assessing health status in mice. Laboratory animal science. 1999;49(3):319–23.

59. Paster E. V, Villines K.A., Hickman D.L. Endpoints for mouse abdominal tumor models: Refinement of current criteria. Comparative Medicine. 2009 Jun;59(3):234–41.

60. Easterly M.E., Foltz C.J., Paulus M.J. Body condition scoring: comparing newly trained scorers and micro-computed tomography imaging. Lab animal. 2001 Mar;30(3):46–9.

61. Noldus L.P.J.J. The Observer: A software system for collection and analysis of observational data. Behavior Research Methods, Instruments, & Computers. 1991 Sep;23(3):415–29.

62. Honma S., Honma K.I., Shirakawa T., Hiroshige T. Rhythms in behaviors, body temperature and plasma corticosterone in SCN lesioned rats given methamphetamine. Physiology and Behavior. 1988 Jan 1;44(2):247–55.

63. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2019.

64. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. Boston, MA; 2015.

65. Maechler M., Rousseeuw P., Struyf A., Hubert M., Hornik K. cluster: Cluster Analysis Basics and Extensions. 2018.

66. Pinheiro J., Bates D., DebRoy S., Sarkar D., R Core Team. {nlme}: Linear and Nonlinear Mixed Effects Models [Internet]. 2018.

67. Brooks M.E., Kristensen K., van Benthem K.J., Magnusson A., Berg C.W., Nielsen A., et al. {glmmTMB} Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal. 2017;9(2):378–400.

68. Valletta J.J., Torney C., Kings M., Thornton A., Madden J. Applications of machine learning in animal behaviour studies. Animal Behaviour. 2017 Feb 1;124:203–20.

69. Cormack J.R., Orme R.M., Costello T.G. The role of α2-agonists in neurosurgery. Journal of Clinical Neuroscience. 2005 May 1;12(4):375–8.

70. Laarakker M.C., Raai J.R. va., van Lith H.A., Ohl F. The role of the alpha 2A-adrenoceptor in mouse stress-coping behaviour. Psychoneuroendocrinology. 2010 May 1;35(4):490–502.

71. Lenth R. emmeans: Estimated Marginal Means, aka Least-Squares Means [Internet]. 2019.

72. Fox J., Weisberg S. An {R} Companion to Applied Regression [Internet]. Second. Thousand Oaks {CA}: Sage; 2011.

73. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016.

74. Abdi H. Bonferroni and Šidák corrections for multiple comparisons. Encyclopedia of measurement and statistics. 2007;3:103–7.

75. Maximino C., van der Staay F.J. Behavioral models in psychopathology: Epistemic and semantic considerations. Behavioral and Brain Functions. 2019 Mar 1;15(1):1.

76. van Der Staay F.J., Arndt S.S., Nordquist R.E. The standardization-generalization dilemma: A way out. Genes, Brain and Behavior. 2010 Nov;9(8):849–55.

77. Huang Q., Zhou Y., Liu L.Y. Effect of post-weaning isolation on anxiety- and depressive-like behaviors of C57BL/6J mice. Experimental Brain Research. 2017 Sep 10;235(9):2893–9.

78. Kwak C., Lee S.H., Kaang B.K. Social isolation selectively increases anxiety in mice without affecting depression-like behavior. Korean Journal of Physiology and Pharmacology. 2009 Oct;13(5):357–60.

79. Carola V., D'Olimpio F., Brunamonti E., Mangia F., Renzi P. Evaluation of the elevated plus-maze and open-field tests for the assessment of anxiety-related behaviour in inbred mice. Behavioural Brain Research. 2002 Aug 21;134(1–2):49–57.

80. Rodgers R.J., Boullier E., Chatzimichalaki P., Cooper G.D., Shorten A. Contrasting phenotypes of C57BL/6JOlaHsd, 129S2/SvHsd and 129/SvEv mice in two exploration-based tests of anxiety-related behaviour. Physiology and Behavior. 2002 Nov 1;77(2–3):301–10.

81. Bolivar V.J. Intrasession and intersession habituation in mice: From inbred strain variability to linkage analysis. Neurobiology of Learning and Memory. 2009 Sep 1;92(2):206–14.

82. Augustsson H., Meyerson B.J. Exploration and risk assessment: A comparative study of male house mice (Mus musculus musculus) and two laboratory strains. Physiology and Behavior. 2004 Jun 1;81(4):685–98.

83. Kulesskaya N., Voikar V. Assessment of mouse anxiety-like behavior in the light-dark box and open-field arena: Role of equipment and procedure. Physiology and Behavior. 2014 Jun 22;133:30–8.

84. van Donkelaar E.L., Vaessen K.R.D., Pawluski J.L., Sierksma A.S., Blokland A., Cañete R., et al. Long-term corticosterone exposure decreases insulin sensitivity and induces depressive-like behaviour in the C57BL/6NCrl mouse. Holscher C, editor. PLoS ONE. 2014 Oct 13;9(10):e106960.

85. Nonogaki K., Nozue K., Oka Y. Social isolation affects the development of obesity and type 2 diabetes in mice. Endocrinology. 2007 Oct;148(10):4658–66.

86. Sun M., Choi E.Y., Magee D.J., Stets C.W., During M.J., Lin E.-J.D. Metabolic Effects of Social Isolation in Adult C57BL/6 Mice. International Scholarly Research Notices. 2014;2014:1–9.

87. Bartolomucci A., Palanza P., Sacerdote P., Ceresini G., Chirieleison A., Panerai A.E., et al. Individual housing induces altered immuno-endocrine responses to psychological stress in male mice. Psychoneuroendocrinology. 2003 May 1;28(4):540–58.

88. Benton D., Goldsmith J.F., Gamal-El-Din L., Brain P.F., Hucklebridge F.H. Adrenal activity in isolated mice and mice of different social status. Physiology and Behavior. 1978 Apr 1;20(4):459–64.

89. An D., Chen W., Yu D.Q., Wang S.W., Yu W.Z., Xu H., et al. Effects of social isolation, re-socialization and age on cognitive and aggressive behaviors of Kunming mice and BALB/c mice. Animal Science Journal. 2017 May 1;88(5):798–806.

90. Brain P.F. What does individual housing mean to a mouse? Life Sciences. 1975 Jan 15;16(2):187–200.

91. Brunner D., Hen R. Insights into the neurobiology of impulsive behavior from serotonin receptor knockout mice. In: Annals of the New York Academy of Sciences. John Wiley & Sons, Ltd (10.1111); 1997. p. 81–105.

92. Popova N.K. From genes to aggressive behavior: The role of serotonergic system. BioEssays. 2006 May 1;28(5):495–503.

93. Frankle W.G., Lombardo I., New A.S., Goodman M., Talbot P.S., Huang Y., et al. Brain serotonin transporter distribution in subjects with impulsive aggressivity: A positron emission study with [11 C]McN 5652. American Journal of Psychiatry. 2005 May;162(5):915–23.

94. Francesco Ferrari P., Palanza P., Parmigiani S., De Almeida R.M.M., Miczek K.A. Serotonin and aggressive behavior in rodents and nonhuman primates: Predispositions and plasticity. European Journal of Pharmacology. 2005 Dec 5;526(1–3):259–73.

95. Brodkin E.S. BALB/c mice: Low sociability and other phenotypes that may be relevant to autism. Behavioural Brain Research. 2007 Jan 10;176(1):53–65.

96. Siuda D., Wu Z., Chen Y., Guo L., Linke M., Zechner U., et al. Social isolation-induced epigenetic changes in midbrain of adult mice. Journal of physiology and pharmacology : an official journal of the Polish Physiological Society. 2014 Apr;65(2):247–55.

97. Sargin D., Oliver D.K., Lambe E.K. Chronic social isolation reduces 5-HT neuronal activity via upregulated SK3 calcium-activated potassium channels. eLife. 2016 Nov 22;5:e21416.

98. Ma J., Wang F., Yang J., Dong Y., Su G., Zhang K., et al. Xiaochaihutang attenuates depressive/anxiety-like behaviors of social isolation-reared mice by regulating monoaminergic system, neurogenesis and BDNF expression. Journal of Ethnopharmacology. 2017 Aug 17;208:94–104.

99. Sallinen J., Link R.E., Haapalinna A., Viitamaa T., Kulatunga M., Sjöholm B., et al. Genetic alteration of alpha 2C-adrenoceptor expression in mice: influence on locomotor, hypothermic,

and neurochemical effects of dexmedetomidine, a subtype-nonselective alpha 2-adrenoceptor agonist. Molecular pharmacology. 1997 Jan;51(1):36–46.

100. Engel O., Masic A., Landsberg G., Brooks M., Mills D.S., Rundfeldt C. Imepitoin shows benzodiazepine-like effects in models of anxiety. Frontiers in Pharmacology. 2018;9(NOV):1225.

101. Naaz S., Ozair E. Dexmedetomidine in current anaesthesia practice- a review. Journal of Clinical and Diagnostic Research. 2014 Oct;8(10):GE01–4.

102. Weerink M.A.S., Struys M.M.R.F., Hannivoort L.N., Barends C.R.M., Absalom A.R., Colin P. Clinical Pharmacokinetics and Pharmacodynamics of Dexmedetomidine. Clinical Pharmacokinetics. 2017;56(8):893–913.

103. Xu B., Li Z., Zhou D., Li L., Li P., Huang H. The influence of age on sensitivity to dexmedetomidine sedation during spinal anesthesia in lower limb orthopedic surgery. Anesthesia and Analgesia. 2017 Dec;125(6):1907–10.

104. Fairbanks C.A., Kitto K.F., Nguyen H.O., Stone L.S., Wilcox G.L. Clonidine and dexmedetomidine produce antinociceptive synergy in mouse spinal cord. Anesthesiology. 2009 Mar 1;110(3):638–47.

105. Wilkinson M., Manchester E.L. Strain differences in brain alpha2 and beta-adrenergic receptor binding in dystrophie mice. Brain Research Bulletin. 1983 Dec 1;11(6):743–5.

106. Rabin B.C., Guo T.Z., Gregg K., Maze M. Role of serotonergic neurotransmission in the hypnotic response to dexmedetomidine, an α2-adrenoceptor agonist. European Journal of Pharmacology. 1996 Jun;306(1–3):51–9.

107. Reinhard J., Roth R. Noradrenergic modulation of serotonin synthesis and metabolism. I. Inhibition by clonidine in vivo. Journal of Pharmacology and Experimental Therapeutics. 1982;221(3):541–6.

108. Gobert A., Rivet J.M., Audinot V., Newman-Tancredi A., Cistarelli L., Millan M.J. Simultaneous quantification of serotonin, dopamine and noradrenaline levels in single frontal cortex dialysates of freely-moving rats reveals a complex pattern of reciprocal auto- and heteroreceptor-mediated control of release. Neuroscience. 1998 Feb;84(2):413–29.

109. Sallinen J., Haapalinna A., Viitamaa T., Kobilka B.K., Scheinin M. D-amphetamine and L-5-hydroxytryptophan-induced behaviours in mice with genetically-altered expression of the α(2C)-adrenergic receptor subtype. Neuroscience. 1998 Jun;86(3):959–65.

110. Rushton W.F., Charlton N.P. Dexmedetomidine in the Treatment of Serotonin Syndrome. Annals of Pharmacotherapy. 2014 Dec 28;48(12):1651–4.

111. Weinger M.B., Chen D.Y., Lin T., Lau C., Koob G.F., Smith N.T. A role for CNS alpha-2 adrenergic receptors in opiate-induced muscle rigidity in the rat. Brain research. 1995 Jan 9;669(1):10–8.

112. Weinger M.B., Segal I.S., Maze M. Dexmedetomidine, acting through central alpha-2 adrenoceptors, prevents opiate-induced muscle rigidity in the rat. Anesthesiology. 1989;71(2):242–9.

113. White W.F., Regan L.J., Roe A.W., Messer A. Behavior, genetics and biochemistry of an allele of the mutant mouse spastic, spalb. Journal of Neurogenetics. 1987 Aug;4(1):253–8.

114. Pariante C.M., Lightman S.L. The HPA axis in major depression: classical theories and new developments. Trends in Neurosciences. 2008 Sep 1;31(9):464–8.

115. Compas B.E. Psychobiological processes of stress and coping: Implications for resilience in children and adolescents - Comments on the papers of Romeo & McEwen and Fisher et al. In: Annals of the New York Academy of Sciences. John Wiley & Sons, Ltd (10.1111); 2006. p. 226–34.

116. Brinks V., van der Mark M., de Kloet R., Oitzl M.S. Emotion and cognition in high and low stress sensitive mouse strains: a combined neuroendocrine and behavioral study in BALB/c and C57BL/6J mice. Frontiers in Behavioral Neuroscience. 2007 Dec 30;1:8.

117. Shanks N., Griffiths J., Zalcman S., Zacharko R.M., Anisman H. Mouse strain differences in plasma corticosterone following uncontrollable footshock. Pharmacology, Biochemistry and Behavior. 1990 Jul;36(3):515–9.

118. Jacobson L., Sapolsky R. The role of the hippocampus in feedback regulation of the hypothalamic-pituitary-adrenocortical axis. Endocrine Reviews. 1991 May 1;12(2):118–34.

119. Herman J.P., Cullinan W.E. Neurocircuitry of stress: Central control of the hypothalamo-pituitary-adrenocortical axis. Trends in Neurosciences. 1997 Feb 1;20(2):78–84.

120. Benedetti M., Merino R., Kusuda R., Ravanelli M.I., Cadetti F., Dos Santos P., et al. Plasma corticosterone levels in mouse models of pain. European Journal of Pain. 2012 Jul 1;16(6):803–15.

121. Anisman H., Hayley S., Kelly O., Borowski T., Merali Z. Psychogenic, neurogenic, and systemic stressor effects on plasma corticosterone and behavior: mouse strain-dependent outcomes. Behavioral Neuroscience. 2001;115(2):443–54.

122. Balcombe J.P., Barnard N.D., Sandusky C. Laboratory routines cause animal stress. Contemporary topics in laboratory animal science. 2004;43(6):42–51.

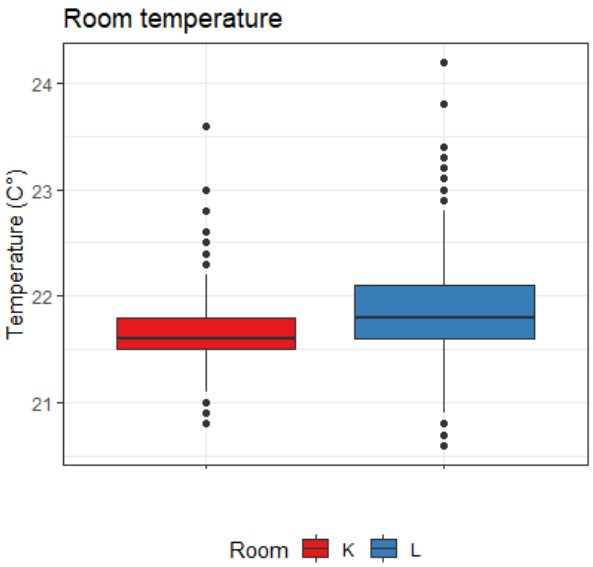
123. Martin T.G., Wintle B.A., Rhodes J.R., Kuhnert P.M., Field S.A., Low-Choy S.J., et al. Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. Ecology Letters. 2005 Nov 1;8(11):1235–46.

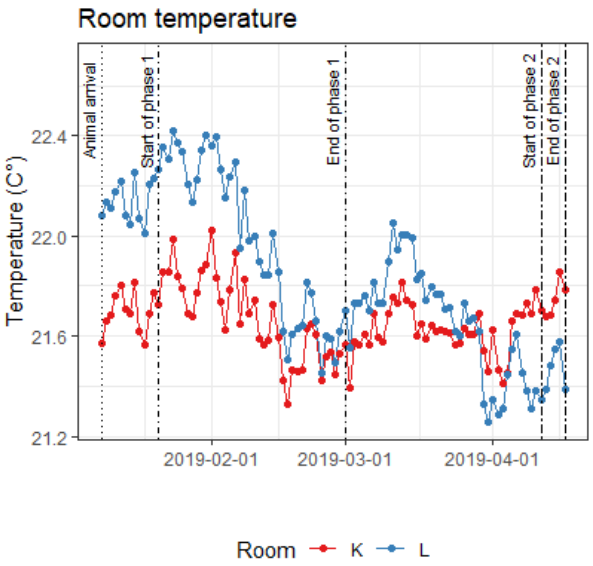
APPENDIX

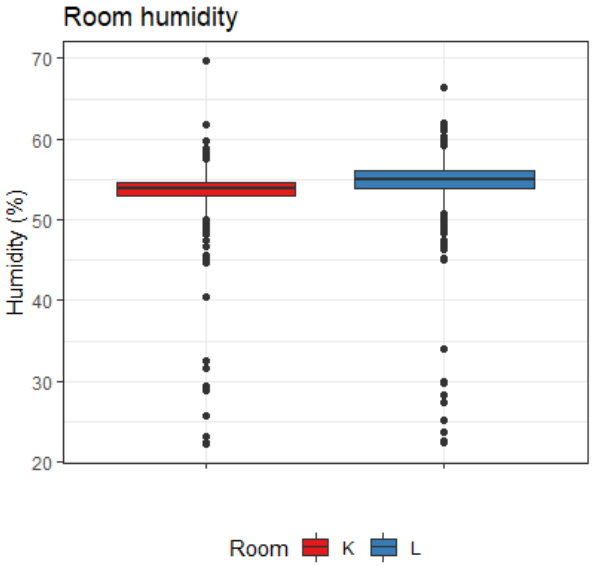
APPENDIX I

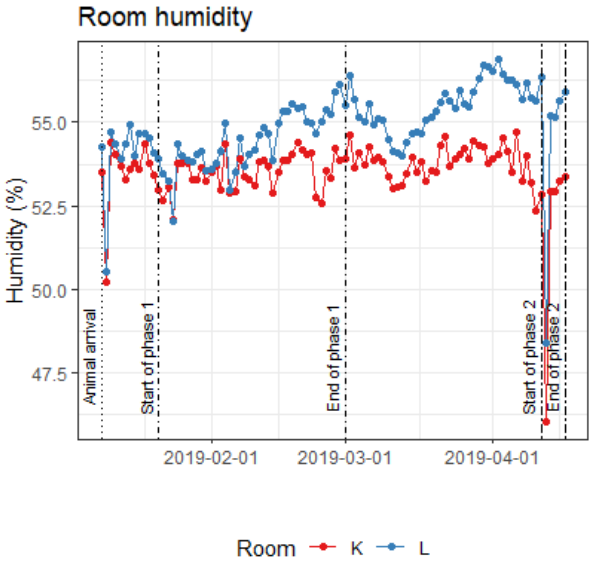
- Avoidance behaviour
    - Latency to initial board visit *(+z)*
    - Frequency of board visits *(-z)*
    - Percentage of time spent on board *(-z)*
- Risk-assessment
    - Frequency of risk-assessments *(+z)*
    - Latency to initial risk-assessment *(-z)*
- Locomotion
    - Frequency of line crossing *(+z)*
    - Latency of immobility *(+z)*
    - Latency to initial line crossing *(-z)*
    - Frequency of immobility *(-z)*
    - Percentage of time spent on immobility *(-z)*
- Exploration
    - Frequency of rearing *(+z)*
    - Frequency of exploration of holes on the board *(+z)*
    - Latency to rearing *(-z)*
    - Latency to exploration of holes on the board *(-z)*
- Arousal
    - Frequency of grooming-bouts *(+z)*
    - Percentage of time grooming *(+z)*
    - Frequency of defecation *(+z)*
    - Latency to grooming-bouts *(-z)*
    - Latency to defecation *(-z)*

APPENDIX II


Room temperature
Temperature (C°)
24
23
22
21
Room
K
L


Room temperature
Temperature (C°)
22.4
22.0
21.6
21.2
Animal arrival
Start of phase 1
End of phase 1
Start of phase 2
End of phase 2
2019-02-01
2019-03-01
2019-04-01
Room
K
L


Room humidity
Humidity (%)
70
60
50
40
30
20
Room
K
L


Room humidity
Humidity (%)
55.0
52.5
50.0
47.5
Animal arrival
Start of phase 1
End of phase 1
Start of phase 2
End of phase 2
2019-02-01
2019-03-01
2019-04-01
Room
K
L