Comparative Severity Assessment of Genetic, Stress-Based, and Pharmacological Mouse Models of Depression

Anne Stephanie Mallien[1]*, Natascha Pfeiffer[1], Christiane Brandwein[1], Dragos Inta[1,2], Rolf Sprengel[3], Rupert Palme[4], Steven R. Talbot[5†] and Peter Gass[1†]

[1] Research Group (RG) Animal Models in Psychiatry, Department of Psychiatry and Psychotherapy, Medical Faculty Mannheim, Central Institute of Mental Health, Heidelberg University, Heidelberg, Germany, [2] Department for Community Health, Faculty of Natural Sciences and Medicine, University of Fribourg, Fribourg, Switzerland, [3] Max Planck Institute for Medical Research (MPIMF), Heidelberg, Germany, [4] Department of Biomedical Sciences, University of Veterinary Medicine, Vienna, Austria, [5] Institute for Laboratory Animal Science, Hannover Medical School, Hanover, Germany

The use of animals in neurosciences is pivotal to gaining insights into complex functions and dysfunctions of behavior. For example, various forms of physical and/or psychological stress are inherent to various animal models for psychiatric disorders, e.g., depression. Regarding animal welfare, it would be mandatory to use models that inflict the least amount of stress necessary to address the underlying scientific question. This study compared the severity of different approaches to induce depression in mice: mutagenesis in GluA1 knockout, immobilization stress, and stress-induction *via* stress hormone treatment. While genetic alterations potentially represent a lifelong burden, the temporary intervention only affects the animals for a limited time. Therefore, we used home cage-based behavioral and physiological parameters, including nest building, burrowing, body weight, and fecal corticosterone metabolites, to determine the well-being of male and female mice. In addition, we performed an evidence-based estimate of severity using a composite score for relative severity assessment (RELSA) with this data. We found that even though restraint stress and supplementation of corticosterone in the diet both aimed at depression-related precipitating stress effects, the latter affected the well-being much stronger, especially in females. Restraint leads to less noticeable well-being impairments but causes depression-associated anhedonic behavior. Mice of both sexes recovered well from the stress treatment. GluA1 KO and their littermates showed diminished well-being, comparable to the immobilization experiments. However, since this is a lifelong condition, this burden is not reversible and potentially accumulative. In line with the 3Rs (Replacement, Reduction, and Refinement), the process of choosing the most suitable model should ideally include an evidence-based severity assessment to be able to opt for the least severe alternative, which still induces the desired effect. Promoting refinement, in our study, this would be the restraint stress.

Keywords: depression model, mouse, severity assessment, Laboratory Animal Science, 3R, GluA1 (AMPA receptor subunit GluR1), stress, fecal corticosterone metabolites

INTRODUCTION

The use of animal models in neurosciences, especially in experimental psychiatry, is pivotal to gaining mechanistic insights into behavior's complex functions and dysfunctions (Homberg et al., 2021). Comprehension of such mechanisms provides the basis for understanding critical aspects of pathogenesis, pathophysiology, and the development of new therapeutic approaches for psychiatric disorders. There is an undeniable necessity for further research in this area. However, since animal experiments and models are mainly based on mimicking the human condition, various forms of physical and/or psychological stress are inherent to a variety of psychiatric animal models, also because stress is an essential component of psychiatric diseases, in particular in affective disorders (Nestler and Hyman, 2010; Neumann et al., 2011).

The use of animals for research is legally regulated on the explicit understanding that such use will provide significant new insights facilitating relevant benefits. No unnecessary harm will be imposed on the animals. Strains and distress should be avoided or minimized. This ethical dilemma – to balance the gain of knowledge with potential harm to the animals – was the basis for the development of the "3R principle" and, more recently, for "harm-benefit-analyses" (Wurbel, 2017; Grimm et al., 2019). Meanwhile, these principles have been integrated into European legislation (Directive 2010/63/EU) and good scientific practice, e.g., in the ARRIVE Guidelines or the Basel Declaration (Kilkenny et al., 2010; Percie du Sert et al., 2020).

No generally accepted standardizing body is available for evidence-based assessment and classification of severity regarding maintenance, handling, and experimental procedures in laboratory animals, particularly rodents (Birnbacher, 2009). Thus, assignment to the severity grades "mild," "moderate," or "severe" in the EU Directive 2010/63 was not evidence – but rather eminence-based, from an anthropomorphic perspective. Rodents might have a different perception due to their species-specific adaptation to the environment. In principle, it is evident that stressors, e.g., foot shocks, social isolation, or restraint, cause considerable distress. But why is a transient state of distress, e.g., learned helplessness, regarded as severe and not as moderate (Mallien et al., 2020)? Do different forms of stress evoke different severity levels? Are behavioral strategies to model depression symptoms more severe than transgenic models with lifelong or induced mutations of candidate genes for a disease? And, Do these stressors, especially if transient, compromise the animals more than experimentally induced cancer or inflammation? These questions are all insufficiently or unanswered in the EU Directive.

In the present study, we used established behavioral and physiological indicators for impaired welfare for a comparative analysis of models for depression in mice. Moreover, our analysis included an algorithm-based comprehensive composite score to detect welfare impairments objectively and provide a relative severity assessment (RELSA) of the different modeling strategies (Talbot et al., 2020). We assessed the severity of several established models for depression based on targeted mutagenesis and stress-based treatments. All models have been shown to induce defined syndromes in mice, leading to alterations in the animals' emotional or cognitive behaviors. *GluA1* knockout mice represent a well-known animal model for depression based on mutation of the AMPA receptor subunit 1 (Zamanillo et al., 1999; Bannerman et al., 2004; Chourbaji et al., 2008; Wiedholz et al., 2008; Sanderson et al., 2009, 2017; Fitzgerald et al., 2010; Inta et al., 2010; Bygrave et al., 2016, 2019; Eltokhi et al., 2020; Ang et al., 2021), which is consistent with the glutamate hypothesis of depression (Sanacora et al., 2012). As stress-based models, we used the daily experience of restraint stress on the one hand and on the other, chronic pharmacological treatment with the stress hormone corticosterone (Moda-Sava et al., 2019). We analyzed and compared the severity before, during, and after the respective stressor/treatment.

MATERIALS AND METHODS

Experimental Design

We aimed to assess models of different modalities: one targeted mutagenesis line in Experiment 1 and two stress-based models in Experiment 2.

In Experiment 1, we measured the severity-related parameters only once since the animals' investigated genetic alterations are life-long conditions (**Figure 1**). In Experiment 2, we observed the severity in the stress-based models in a longitudinal approach to detect the exacerbation due to the treatment and the subsequent alleviation afterward (**Figure 1**). The experimenters were blinded for genotype and treatment during the behavioral experiments.

Animals

In Experiment 1, we investigated the genetically modified mouse line *Gria1$^{-/-}$* (Zamanillo et al., 1999). The *Gria1* gene encodes the GluA1 subunit of the AMPA receptor. GluA1 KO mice were bred in the Interfacultary Biomedical Faculty (IBF, Heidelberg, Germany). We tested 11 male and 14 female GluA1 knockout mice and the respective 14 male and 11 female littermate controls in 2 cohorts. The mice were 7–10 weeks old at arrival at the Central Institute of Mental Health (CIMH, Mannheim, Germany).

For Experiment 2, we purchased 8-week-old male and female C57BL/6N mice from Charles River (Charles River Laboratories, Sulzfeld, Germany) and assigned 10 per sex to 3 different groups: *restrained*, *corticosterone*, and *control*. Saccharin preference (SP) and burrowing performances were used to distribute the mice into groups with equivalent baseline behaviors pseudorandomly.

All animals were single-housed in Type II cages with bedding (Abedd Espen MIDI, ABEDD, Vienna, Austria), cotton nestlet (Zoonlab, Castrop-Rauxel, Germany), and *ad libitum* access to food (LasQCdiet Rod16-H, Altromin, Soest, Germany) and tap water. The housings were temperature- ($22 \pm 1°C$) and humidity-controlled ($45 \pm 5\%$) under 12:12 h dark:light cycle with lights off at 7 a.m. All procedures were approved by the German animal welfare authorities (35-9185-81-G-198-17, Regierungspräsidium Karlsruhe) and performed strictly according to the regulations of animal experimentation within the European Union (European Communities Council Directive 2010/63/EU).

Treatment

For the chronic corticosterone treatment (Moda-Sava et al., 2019), we diluted corticosterone in 100% ethanol and added water to receive a final concentration of 0.1 mg/ml corticosterone and 1% ethanol. This was supplied to the animals in their drinking bottle. Since two female mice reached the humane endpoint and were euthanized after 18 days, we decided to shorten the chronic treatment from the initially planned 21 days to only 18.

We adjusted the duration of repeated restraint accordingly, which was then performed for 18 days with physical restriction for 1 h per day in plastic tubes (26 mm diameter) with holes for breathing in a separate experimental room during the dark phase.

Behavioral Analysis

Nesting

Nest building was evaluated according to a rating scale on the shape and cohesion of the nest as previously described (Deacon, 2006; Mallien et al., 2020). The mice were placed in a new home cage with a cotton nestlet 1 h before the onset of the dark phase, and the score was determined after 13 h. Subsequently, we performed the "Time to integrate into nest test" (TINT). For this, additional nesting material (sizzle, Zoonlab, Castrop-Rauxel, Germany) was introduced into the diagonally opposing corner of the nest site within the first 3 h of the light phase, and the latency to integrate the novel material into the existing nest site was observed for 10 min (Hager et al., 2015; Mallien et al., 2020). In the GluA1 study, we also performed a daily nest scoring, always at the same time of day [corresponding to the 13 h timing of the nest test (NT)].

Burrowing

To analyze the burrowing behavior, we placed bottles (14 cm long × 5.5 cm in diameter) filled with food pellets at the rear of the home cage 1 h before the dark phase and let the mice burrow as previously described (Sensini et al., 2020). The amount of burrowed material (% of total weight) was assessed after 6 h. All mice were accustomed to the procedure for five consecutive days before the final testing day.

Fecal Corticosterone Metabolites

Feces samples were collected for 24 h in a secondary home cage and were processed as described (Sensini et al., 2020). Briefly, an extract of dried and homogenized feces was produced with 80% methanol, and an aliquot was analyzed in a well-established and validated 5α-pregnane-3β,11β,21-triol-20-one enzyme immunoassay (EIA) (Touma et al., 2003, 2004; Palme, 2019; Sensini et al., 2020).

Saccharin Preference

We investigated the preference for saccharin solution of the stress-based models as previously described (Kronenberg et al., 2012; Klein et al., 2015; Boldt et al., 2021). Briefly, we determined the intake from two drinking bottles in 24 h on four consecutive days with either two bottles filled with water (day 1 and 3) or one with water and one with a 0.1% saccharin solution (day 2 and 4). Day 1 and 3 were used to detect putative side bias. On days 2 and 4, we determined the preference for saccharin.

Statistical Analyses

We analyzed all outcome measures [nesting, burrowing, fecal corticosterone metabolites (FCMs), SP] with linear models. In Experiment 1, the dependent variables treatment, cohort, and sex were modeled as interactions (parameter ∼ treatment: cohort: sex). In Experiment 2, the dependent variables treatment, phase, and sex were modeled as interactions (parameter ∼ treatment: phase: sex). Factors were treated as simple treatment contrasts with the default level in the intercept. Subsequent multiple comparisons were tested with the Tukey–Kramer *post hoc* test. The loss of two highly burdened females might have distorted the results of the post-stress phase for the females. No animals were excluded from the study or any statistical analyses. The single animal served as an experimental unit.

We used the RELSA (Talbot et al., 2020) score to compare different models. This algorithm-based comprehensive composite score detects relative welfare impairments from multi-dimensional input parameters of physiology and behavior. We chose to include body weight (BW), burrowing performance, and fecal corticosterone metabolite concentrations as the most suitable parameters to determine the severity. RELSA achieved relative comparability by expressing the multi-dimensional change in the input variables' differences regarding a reference set with a fixed severity quality. In this study, the reference set was obtained from values in the restrained experiment. In addition, the time-independent maximum RELSA score ($RELSA_{max}$) in each animal was used to compare different experiments and subgroups quantitatively. Finally, we used the RELSA to present the state of severity over time ($RELSA_{flow}$) or the most pronounced severity displayed by the animals throughout the procedure ($RELSA_{max}$).
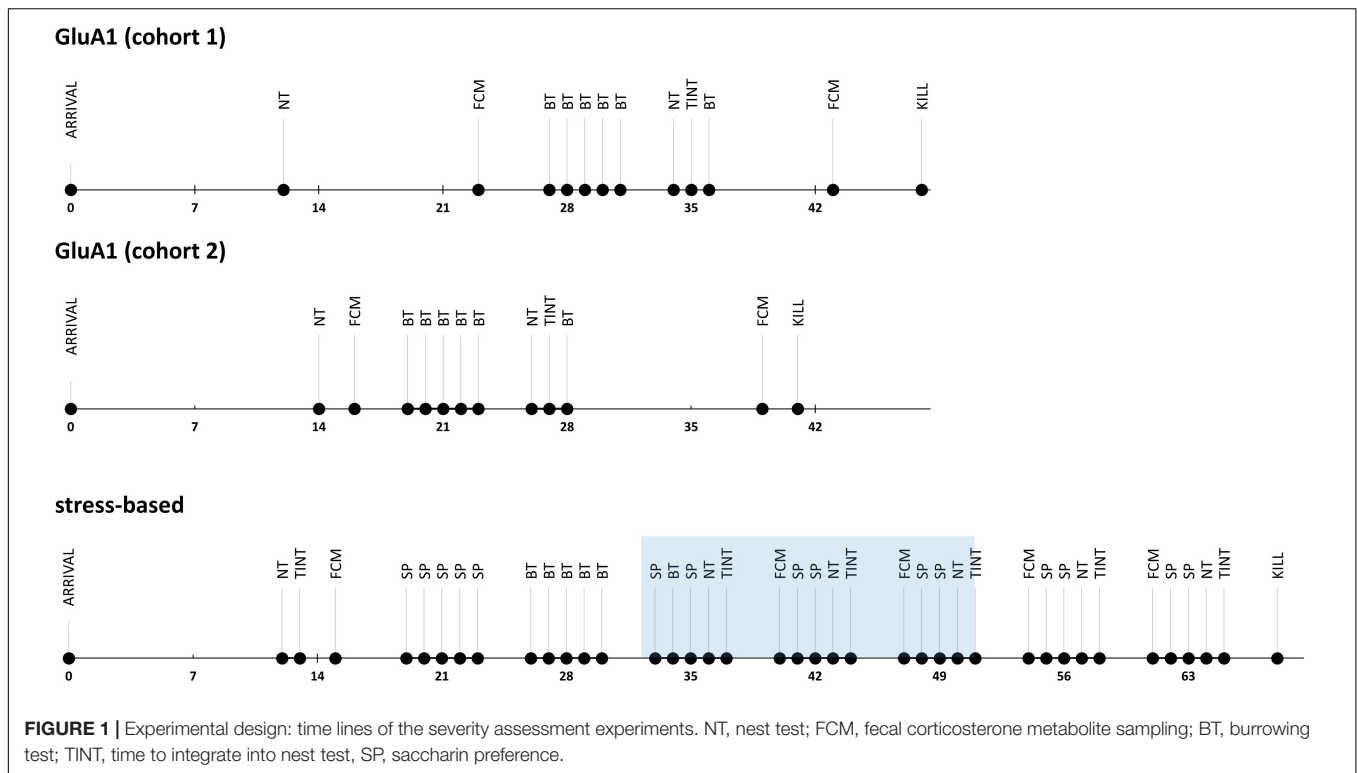
RESULTS

Experiment 1: Severity Assessment of GluA1 KO Mice

Body Weight

GluA1 KO and WT mice showed normal weight gain over time [time: $F(1,234) = 109.98$, $p < 0.0001$] and the typical sex differences with females weighing 5.19 (SE 0.54) g less than the intercept [β = 26.23 (SE 0.31) g, $p < 0.0001$] (**Figure 2A**). The mean weight of GluA1 KO was significantly reduced by 1.92 (SE 0.54) g ($p = 0.0004$). This reduction was mainly driven by differences between the male KO and WT mice [cohort I: male GluA1 WT – male GluA1 KO: β = 1.917 (SE 0.539) g, $p = 0.0025$; cohort II: male GluA1 WT – male GluA1 KO: β = 2.564 (SE 0.490) g, $p < 0.0001$], while no differences in BW became apparent for females of the different genotypes. This also led to a treatment: sex interaction [$F(1,234) = 43.19$, $p < 0.0001$]. No differences were detected between the cohorts.

Burrowing Test

The overall estimate of burrowed material in Experiment 1 was 45.09% (SE 4.19) (=intercept) (**Figure 2B**). The GluA1 KO led to a decrease of 39.25% (SE 7.26) ($p < 0.0001$) of burrowed material.

FIGURE 1 | Experimental design: time lines of the severity assessment experiments. NT, nest test; FCM, fecal corticosterone metabolite sampling; BT, burrowing test; TINT, time to integrate into nest test, SP, saccharin preference.

Also, female mice burrowed 32.04% (SE 7.26) less. Female mice burrowed almost not at all and showed no significant differences due to treatment. The female GluA1 KO burrowed 7.84% (SE 4.19) and 15.81% (SE 4.84) (in cohorts I and II, respectively), while their controls burrowed 13.05% (SE 5.93) and 28.83% (SE 5.93). The maximum performance was detected in male controls with 45.09% (SE 4.19) and 29.52% (SE 4.84). The burrowing performance was higher in cohort II [cohort: $F(1,252) = 5.4911$, $p = 0.0199$], driven by the relatively high performance of male GluA1 KO mice with 37.95% (SE 4.48) estimated marginal means. The other GluA1 KO burrowed less: males cohort I 5.874% (SE 5.93); females cohort I 7.84% (SE 4.19); females cohort II 15.81% (SE 4.84).

Nest Test
The GluA1 KO mice built less complex nests with 1.61 (SE 0.50) ($p = 0.0017$) lower complexity scores compared to the intercept [$\beta = 3.73$ (SE 0.29), $p < 0.0001$]. No general sex-specific or cohort-dependent differences were found (**Figure 2C**). Nests of females scored 0.93 (SE 0.50) ($p = 0.0505$) less. Differences between male mice became obvious in the direct comparison: the KO scored lower 1.61 (SE 0.50) ($p = 0.0089$) in cohort I and 1.81 (SE 0.45) ($p = 0.0006$) in cohort II. In female mice, there were no differences between KO and WT performance in cohort I, but in cohort II the score of KO was 2.43 (SE 0.45) ($p < 0.0001$) lower.

Time to Integrate Into Nest Test
We found a general delay in time to integrate the material into the nest [genotype: $F(1,26) = 10.6566$, $p = 0.003069$], indicating a more hesitant behavior in GluA1 KO mice [253.75 (SE 122.42)
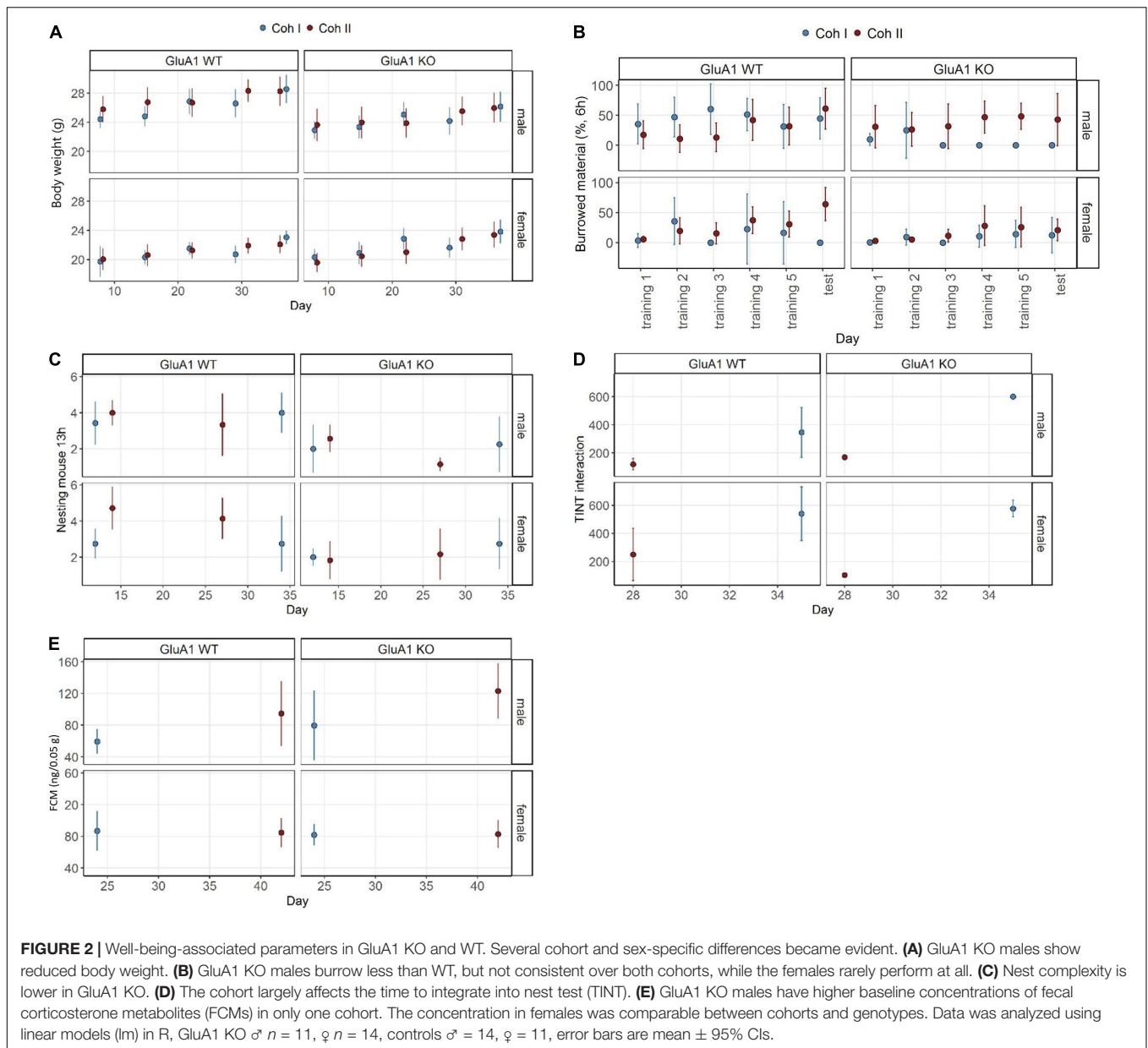
s, $p = 0.0482$] compared to the intercept [$\beta = 346.25$ (SE 54.75) s, $p < 0.0001$] (**Figure 2D**). Female mice generally showed higher latencies [193.75 (SE 94.83) s, $p = 0.0513$]. However, we could not find severe differences in direct comparisons of estimated marginal means between the groups within the same cohort and same sex. There was an overall effect of sex [(sex: $F(1,26) = 5.07$, $p = 0.0330$], but similarly, no significant differences in direct comparisons.

Fecal Corticosterone Metabolite Sampling
Fecal corticosterone metabolite concentration of GluA1 KO mice tended to be increased [$F(1,42) = 3.95$, $p = 0.0534$] (**Figure 2E**). The intercept of FCM was 59.4 (SE 8.8) ng/0.05 g and was lower in GluA1 KO mice [20.2 (SE 15.2) ng/0.05 g, $p = 0.1920$]. In cohort I, the concentrations were significantly higher [63.6 (SE 12.9) ng/0.05 g, $p = 0.0001$] in male KO than in controls. This was neither found in male mice of cohort II nor in genotype comparisons between female mice. No sex effect was evident, but a cohort effect [cohort: $F(1,62) = 9.81$, $p < 0.0032$] was observed. The concentrations of cohort II were significantly decreased [$\beta = 35.1$ (SE 13.4) ng/0.05 g, $p = 0.0122$].

Experiment 2: The Development of Severity in Stress-Based Models Over Time in Restrained, Corticosterone, and Control Mice
We separated the data into three phases for this analysis: the initial baseline assessment, the stress phase during the treatment, and the post-stress evaluation (**Figure 3**).
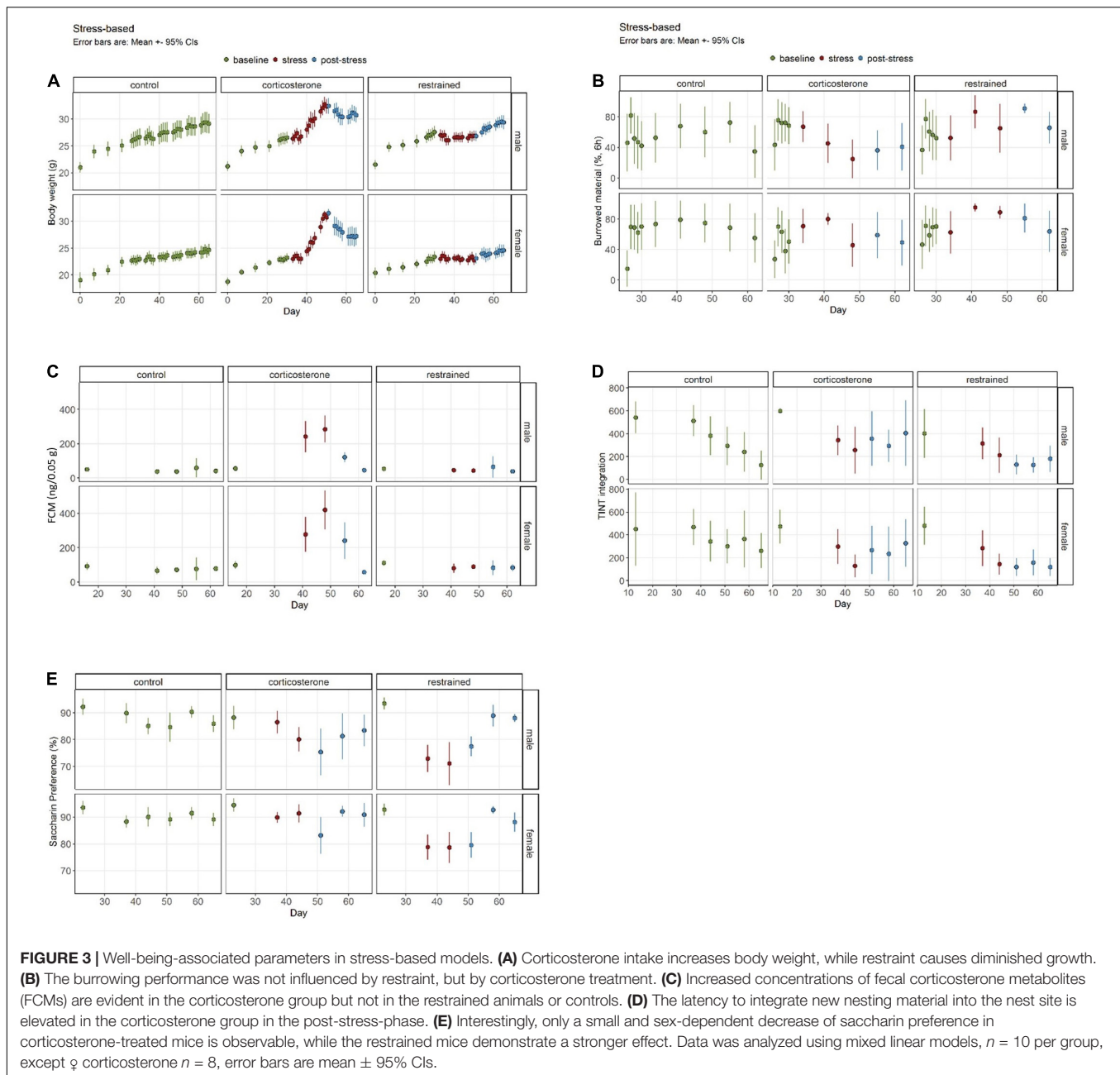
FIGURE 2 | Well-being-associated parameters in GluA1 KO and WT. Several cohort and sex-specific differences became evident. (A) GluA1 KO males show reduced body weight. (B) GluA1 KO males burrow less than WT, but not consistent over both cohorts, while the females rarely perform at all. (C) Nest complexity is lower in GluA1 KO. (D) The cohort largely affects the time to integrate into nest test (TINT). (E) GluA1 KO males have higher baseline concentrations of fecal corticosterone metabolites (FCMs) in only one cohort. The concentration in females was comparable between cohorts and genotypes. Data was analyzed using linear models (lm) in R, GluA1 KO ♂ n = 11, ♀ n = 14, controls ♂ = 14, ♀ = 11, error bars are mean ± 95% CIs.

Body Weight

For the BW we found the typical increase of weight over time [phase: $F_{(2,1919)}$ = 941.21, $p < 0.0001$] and the sex differences [sex: $F_{(1,1919)}$ = 2717.93, $p < 0.0001$] (Figure 3A). The intercept was 66.4 (SE 5.1) g ($p < 0.0001$) and was 0.7 (0.3) g ($p = 0.0251$) increased in restrained mice. Females weighed 3.2 (0.3) g ($p < 0.0001$) less than the intercept. In the males, we found no significant differences in the estimated marginal means during the baseline phase between the restrained group with 25.9 (SE 0.2) g and the corticosterone group with 25.2 (SE 0.2) g. During the stress phase, the mean BW of *corticosterone* mice increased significantly by 4.0 (SE 0.3) g ($p < 0.0001$). The mean BW of *restrained* male mice barely increased by only 0.8 (SE 0.3) g ($p = 0.0583$). In the post-stress phase, both

male *corticosterone* mice and *restrained* male mice significantly gained a comparable amount of weight [*corticosterone*: 1.9 (SE 0.3) g, $p < 0.0001$; *restrained*: 1.9 (SE 0.3) g, $p < 0.0001$]. Overall, the treatments led to significant differences in males ($p < 0.0001$) during the stress phase, with higher BWs in the corticosterone-treated group of 2.5 (SE 0.3) g. This effect was still pronounced in the post-stress phase with 2.5 (SE 0.3) g ($p < 0.0001$).

The estimated marginal means of BW of female mice in the baseline condition was similar for both treatment groups with *corticosterone* 22.0 (SE 0.2) g and *restrained* 22.2 g (SE 0.2) g ($p = 0.9709$). Corticosterone treatment in female mice led to a pronounced increase of 4.1 (SE 0.3) g ($p = 0.0001$) and a smaller increase in the restrained group of 0.9 (SE 0.3) g ($p = 0.0217$)

FIGURE 3 | Well-being-associated parameters in stress-based models. **(A)** Corticosterone intake increases body weight, while restraint causes diminished growth. **(B)** The burrowing performance was not influenced by restraint, but by corticosterone treatment. **(C)** Increased concentrations of fecal corticosterone metabolites (FCMs) are evident in the corticosterone group but not in the restrained animals or controls. **(D)** The latency to integrate new nesting material into the nest site is elevated in the corticosterone group in the post-stress-phase. **(E)** Interestingly, only a small and sex-dependent decrease of saccharin preference in corticosterone-treated mice is observable, while the restrained mice demonstrate a stronger effect. Data was analyzed using mixed linear models, $n = 10$ per group, except ♀ corticosterone $n = 8$, error bars are mean ± 95% CIs.
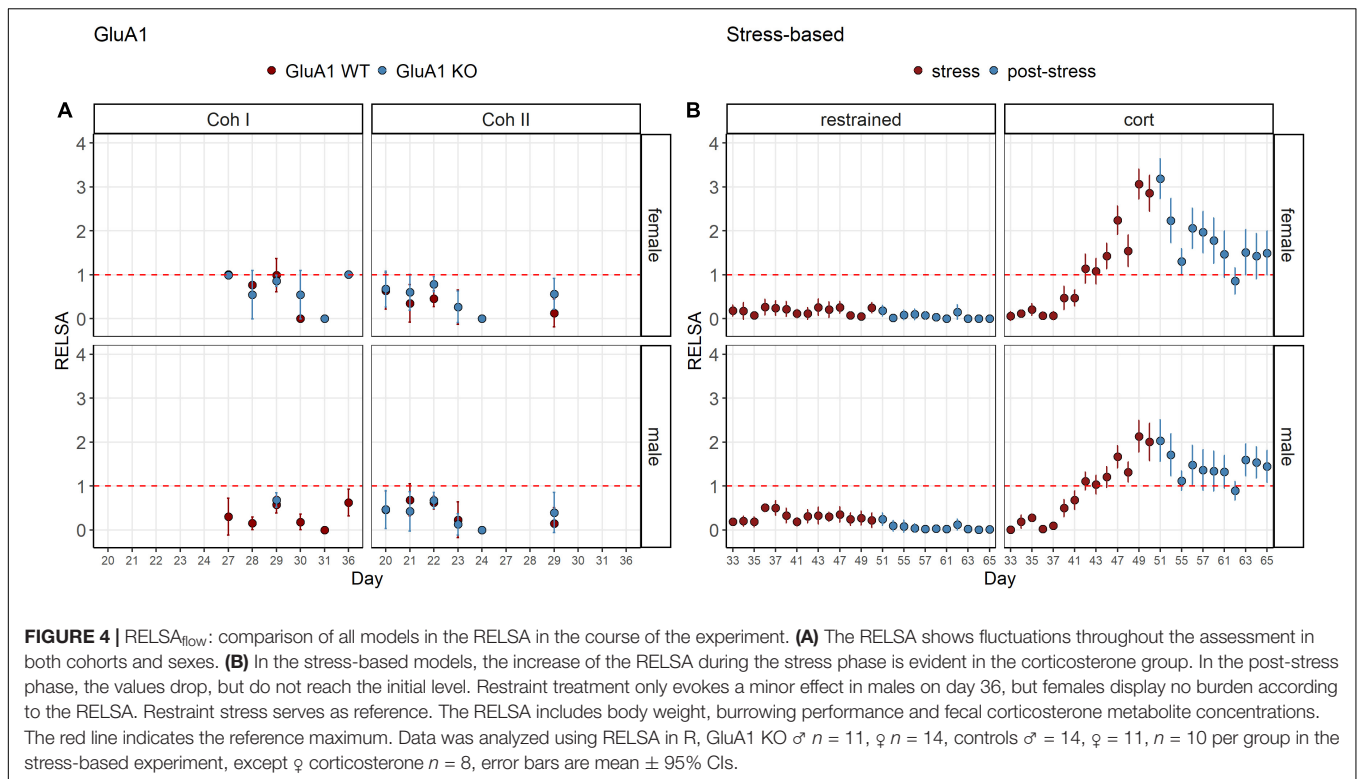
in the stress phase. During the post-stress phase, the weight of the corticosterone group was significantly higher than in the stress phase with 2.2 g (SE 0.3) g ($p < 0.0001$) and for the restrained with 1.0 (SE 0.3) g ($p = 0.0036$). When comparing the different treatments in the stress phase, the female *corticosterone* group weighed 2.9 g (SE 0.3) ($p < 0.0001$) more than the female *restrained* and in the final post-stress phase, the difference was even more prominent with 4.2 g (SE 0.3 g) ($p < 0.0001$).

In comparison, males gained 15.9% and females 18.2% of total BW during corticosterone treatment in the stress phase. This indicates a higher sensitivity toward corticosterone treatment in female mice.

Burrowing Test

The overall estimate of burrowed material in Experiment 1 was 66.4% (SE 5.1%) (=intercept) (**Figure 3B**). We found a general difference due to treatment [treatment: $F(1,383) = 11.7301$, $p = 0.0007$] in the performance of the burrowing behavior. No general alterations due to sex or phase were detected. However, performance dropped significantly in the male *corticosterone* group from baseline 66.4% (SE 5.1) to 38.4% (SE 8.1) in the post-stress phase ($p = 0.0441$). All other groups showed no significant effects due to the phase.

In the baseline phase, no treatment effects were observed within the same sex. However, in the stress phase, male

FIGURE 4 | RELSA_flow: comparison of all models in the RELSA in the course of the experiment. **(A)** The RELSA shows fluctuations throughout the assessment in both cohorts and sexes. **(B)** In the stress-based models, the increase of the RELSA during the stress phase is evident in the corticosterone group. In the post-stress phase, the values drop, but do not reach the initial level. Restraint treatment only evokes a minor effect in males on day 36, but females display no burden according to the RELSA. Restraint stress serves as reference. The RELSA includes body weight, burrowing performance and fecal corticosterone metabolite concentrations. The red line indicates the reference maximum. Data was analyzed using RELSA in R, GluA1 KO ♂ $n = 11$, ♀ $n = 14$, controls ♂ $= 14$, ♀ $= 11$, $n = 10$ per group in the stress-based experiment, except ♀ corticosterone $n = 8$, error bars are mean ± 95% CIs.

corticosterone burrowed 22.26% (SE 9.39) less than the *restrained* group ($p = 0.0846$). This effect was even more pronounced in the post-stress phase when the male *corticosterone* burrowed 40.00% (SE 11.50) less than the restrained ($p = 0.0032$). No effects became apparent for female comparisons of the treatment groups in all phases.

Fecal Corticosterone Metabolite Sampling

We observed the expected increase of FCM concentration in the *corticosterone*-treated mice (**Figure 3C**). The intercept was 56.2 (SE 25.0) ng/0.05 g. The concentration increased by 209.7 (SE 31.2) ng/0.05 g ($p < 0.0001$) compared to baseline 56.2 (SE 25.0) ng/0.05 g in male mice and 293.4 (SE 30.6) ng/0.05 g ($p < 0.0001$) in female mice [baseline phase 98.6 (SE 25.0) ng/0.05 g]. In the post-stress phase values normalized and decreased by 182.0 (SE 25.7) ng/0.05 g ($p < 0.0001$) and 200.2 (SE 26.5) ng/0.05 g ($p < 0.0001$), respectively. Post-stress values were not significantly different to baseline. Restraining did not alter FCM concentrations in our study.

Nest Test

Nesting scores were unaltered throughout the experiments (data not shown).
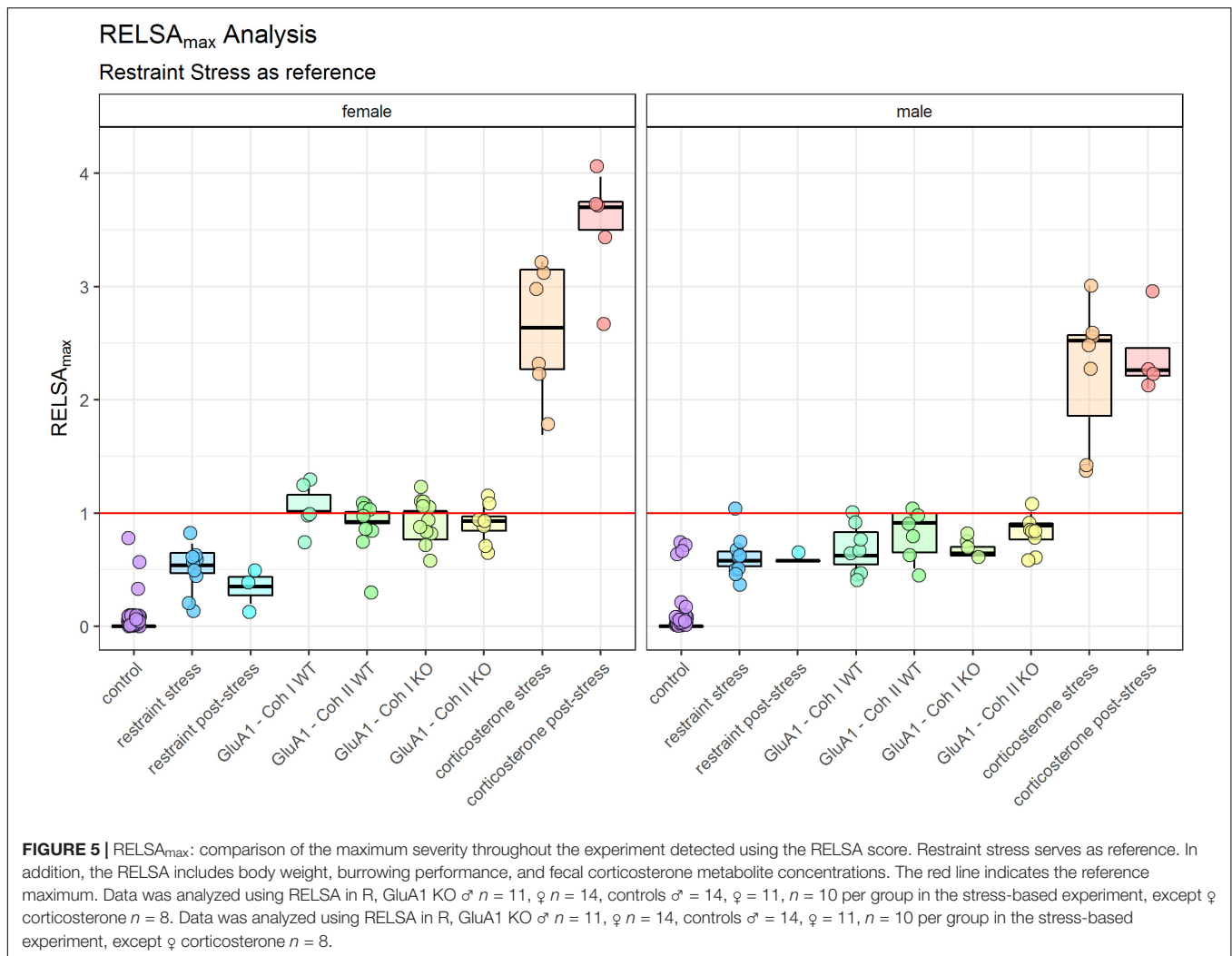
Time to Integrate Into Nest Test

The intercept for the latency to integrate the material into the nest site was 600.0 (SE 72.2) s ($p < 0.0001$). This was influenced by the treatment [$F_{(1,206)} = 16.75$, $p < 0.0001$]. Overall, *restrained* mice were 197.3 (SE 102.1) s faster than the intercept ($p = 0.055$). We also found effects for phase

[$F_{(2,206)} = 25.8974$, $p < 0.0001$] and a treatment: phase interaction [$F_{(2,206)} = 3.64$, $p = 0.0278$]. Again, no differences between the groups were found in the baseline phase. Neither did the treatment groups differ in the stress phase. We did see a general drop in latencies during the experiment. However, that was also visible in the control group. Additionally, in the post-stress phase, the latencies were significantly increased in the *corticosterone* groups in both male mice [196.6 (SE 52.9) s higher than *restrained*, $p = 0.0015$] and female mice [145.9 s (SE 52.9) higher than *restrained*, $p = 0.0321$]. Male mice tended to show higher latencies than females [sex: $F_{(1,206)} = 3.05$, $p = 0.0823$].

Saccharin Preference

The mice showed altered preferences to saccharin drinking due to the treatment [treatment: $F_{(1,220)} = 7.28$, $p = 0.0074$] and also in different experimental phases [phase: $F_{(2,220)} = 29.13$, $p < 0.0001$], resulting in a significant interaction [treatment × phase $F_{(2,220)} = 21.3853$, $p = 3.259e{-}09$] (**Figure 3E**). Overall the intercept was 88.2081% (SE 2.3136). Before the treatment, all groups showed similar preferences, but within the progression of the experiment, the preference developed differently for each treatment and sex [sex: $F_{(1,220)} = 33.30$, $p < 0.0001$]. While there was a strong effect in *restrained* mice, the corticosterone treatment evoked no or only a mild effect. Male *corticosterone* mice showed no significant differences, neither between the baseline and the stress phase nor between the stress an the post-stress phase. Instead, an overall decrease from baseline [88.2% (SE 2.31) to post-stress (80.0% (SE 1.34) of 8.2%

FIGURE 5 | RELSA$_{max}$: comparison of the maximum severity throughout the experiment detected using the RELSA score. Restraint stress serves as reference. In addition, the RELSA includes body weight, burrowing performance, and fecal corticosterone metabolite concentrations. The red line indicates the reference maximum. Data was analyzed using RELSA in R, GluA1 KO ♂ $n = 11$, ♀ $n = 14$, controls ♂ $= 14$, ♀ $= 11$, $n = 10$ per group in the stress-based experiment, except ♀ corticosterone $n = 8$. Data was analyzed using RELSA in R, GluA1 KO ♂ $n = 11$, ♀ $n = 14$, controls ♂ $= 14$, ♀ $= 11$, $n = 10$ per group in the stress-based experiment, except ♀ corticosterone $n = 8$.

(SE 2.67), $p = 0.0295$] was visible. The female *corticosterone* mice showed very stable SP without differences between any phases. On the other hand, the *restrained* mice were highly affected by the treatment: male mice dropped from initially 93.5% (SE 2.31) by 21.43% (SE 2.83) ($p < 0.0001$) in the stress phase, female mice dropped from 92.8% (SE 2.31) by 14.0% (SE 2.83) ($p < 0.0001$). This was 11.3% (SE 2.31) and 11.9% (SE 2.31) less than in the male and female corticosterone groups. As soon as the treatment stopped, the preference for saccharin returned. The females showed no difference between baseline and post-stress preference. In males, the recovery was not as strong. Although the preference increased by 8.04% (SE 2.11) ($p < 0.001$) after the stress, the overall preference was only 84.8% (SE 1.34) and hence lower than before the treatment ($p = 0.0165$).

The RELSA$_{flow}$ showed the severity over time (**Figure 4**). Although no stressful interventions were performed on the GluA1 KO mice, the severity was not as stable as in the *restrained* animals. Those showed nearly no effect even in the

stress phase. There was only a very mild elevation on day 36. On the other hand, the *corticosterone* group revealed an increase that climaxed at the treatment's end and then dropped. However, the RELSA score did not recover to the previous baseline on day 33. Still, it remained on a plateau above the score of 1, which equals the maximum severity of the restrained group.

The maximum severity throughout the experiment for each group is displayed in the RELSA$_{max}$ (**Figure 5**). We found significant differences between male and female mice (**Table 1**). Even though the controls were only housed and received no specific treatment, some still displayed a deflection on the RELSA score at some time in the experiment. But in total, it was evident that this group served as a good baseline of severity. All other groups showed significantly higher RELSA scores – except the male mice in the post-restraint phase. We found more effects in female mice between the different models: GluA1 had higher scores than the *restrained* and the *corticosterone* groups higher scores than the *restrained* and GluA1 group. In males, we found no differences between

TABLE 1 | Results of statistical results from RELSA$_{max}$.

Kruskal–Wallis

	Chi-squared	df	p
Females	241.8	8	<2.2e−16
Males	99.101	8	<2.2e−16

Pairwise comparison (Wilcox-test, Holm corrected)

		p			
		Females		**Males**	
Control	Restraint stress	1.20E−33	****	4.20E−09	****
Control	Restraint post-stress	1.20E−21	****	n.s.	
Control	GluA1 – Coh I WT	1.40E−28	****	7.40E−09	****
Control	GluA1 – Coh II WT	1.50E−34	****	5.10E−08	****
Control	GluA1 – Coh I KO	8.90E−37	****	1.40E−05	****
Control	GluA1 – Coh II KO	2.90E−32	****	1.30E−09	****
Control	Corticosterone stress	1.30E−30	****	1.90E−06	****
Control	Corticosterone post-stress	1.40E−28	****	1.30E−06	****
Restraint stress	GluA1 – Coh I KO	6.60E−03	**	n.s.	
Restraint stress	GluA1 – Coh II KO	3.10E−02	*	n.s.	
Restraint stress	Corticosterone stress	4.20E−02	*	2.80E−01	*
GluA1 – Coh I WT	Corticosterone stress	n.s.		9.00E−03	**
GluA1 – Coh II WT	Corticosterone stress	4.20E−02	*	n.s.	
GluA1 – Coh I KO	Corticosterone stress	2.50E−02	*	n.s.	
GluA1 – Coh II KO	Corticosterone post-stress	4.40E−02	*	0.038	*

*$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$.

GluA1 and *restrained*, but an elevation of *corticosterone* in the GluA1 groups.

DISCUSSION

This study aimed to determine the severity of several standard models for depression in mice and directly compare the relative burden introduced by different strategies, namely targeted mutagenesis of the depression-associated *Gria1* gene, stress-induction *via* subchronic oral corticosterone treatment, or temporary repetitive restraint. This is especially interesting as the various models will affect the subjects for different lifespans: while the GluA1 KO is lifelong and already relevant during development, the stress-based models can vary in duration and the animal's age during onset. Since most depression-based animal experiments are performed in early adulthood, we compared the severity in this phase.

We assume that animals in depression models are affected negatively when compared to controls, even in a home cage-based analysis. As the strategies to induce depression in the animals were created to mimic the symptoms of depressive patients, it is plausible to assume that the well-being in such a model would be as impaired as it is in humans. Yet, the evaluation of suffering should be based on the animals' state rather than anthropomorphic predictions. Hence, the severity assessment was based on established empirical measures of physiology and

behavior (Bleich and Tolba, 2017; Moller et al., 2018; Koska et al., 2019; Seiffert et al., 2019; Bleich et al., 2020; Jirkof et al., 2020; van Dijk et al., 2020; Boldt et al., 2021; Buchecker et al., 2022). Furthermore, we have chosen to use parameters that hardly interfere with the animals and, therefore, should not confound the results by additional stress, e.g., more elaborate behavioral tests like tests for anxiety-related behavior. The behavioral parameters were derived from the natural behavioral repertoire of the mice in the home cage. In addition, these tests are all straightforward to perform, minimize the direct contact with the investigator and can be used by anybody who wants to assess the severity of their setups. In principle, this kind of evaluation is also possible simultaneously with experimentation due to the non-invasive nature of the tests.

GluA1 KO Animals Show Impairment of Well-Being

We started our assessment with GluA1 KO mice and their littermate controls. GluA1 is a subunit of the AMPA subtype of glutamate receptors. Disturbances in the glutamate system are associated with psychiatric disorders, including mood disorders and schizophrenia (Pittenger et al., 2007; Sanacora et al., 2012; Eltokhi et al., 2020). Many studies have thoroughly described the mouse line used in this project. The loss of GluA1 leads to an increase in depression-associated learned helplessness (Chourbaji et al., 2008), hyperlocomotion in

the open field (Wiedholz et al., 2008), novelty- and stress-induced locomotor hyperactivity, altered coping in the forced swim test and in approach-avoidance conflict tests (Fitzgerald et al., 2010), cognitive impairments in short-term memory, puzzle-solving and attention (Zamanillo et al., 1999; Bannerman et al., 2004; Sanderson et al., 2009; Ben Abdallah et al., 2011; Austen et al., 2021; Strickland et al., 2021). In addition, they show circadian disturbances, striking rest-activity pattern changes, and even reduced voluntary running (Ang et al., 2021), another valuable home cage-based parameter for well-being (Mallien et al., 2020; Weegh et al., 2020, 2021).

Some, but not all of these studies, also included female mice. We aimed to assess baseline severity in both sexes. We found the typical BW reduction in male GluA1 KO, not observed in females (Bannerman et al., 2004). The natural behavioral traits were compromised in GluA1 KO: the mice burrowed less, built less complex nests, and were slower in integrating new nesting material to their nest site. This indicates an impairment in well-being, which was cohort-dependent to some degree. One would assume that the same genotype would always create a similar burden on the animals. Instead, we found some effects only to be relevant for one cohort but not the other, sometimes even in a sex-dependent manner, e.g., the nesting of GluA1 KO females was only impaired in cohort II.

In contrast, nesting was impaired in both cohorts of males. This might be triggered by environmental factors or other differences in husbandry (Wurbel, 2017; Jirkof et al., 2020). FCM concentrations of baseline housing were comparable between KO and WT. Overall, one can see that the males' and females' burden is similar in the RELSA$_{flow}$ (**Figure 4**) and RELSA$_{max}$ (**Figure 5**). Moreover, although we found differences for some parameters, there is no observable difference between the KO and the WT in the RELSA.

The SP in GluA1 KO mice has been discussed previously and was not implemented in our current study. Austen et al. (2017) showed that GluA1 KO mice have a normal sucrose consumption level but reduced licking rates. Maksimovic et al. (2014) on the other hand, found increased sucrose preference compared to wild-type controls. As GluA1 KO mice often show hyperlocomotion, this preference might also be due to the need to compensate high energy. Strickland et al. (2021) even question whether GluA1 is implicated in the hedonic value at all.

The poor burrowing performance primarily drives the higher severity scores. Although mice were shown to consistently and vigorously burrow food pellets in the first 2 h of exposure (Deacon, 2009), we observed that the performance was delayed in our laboratory. Hence, we decided to observe the 6-h burrowing interval instead. In other studies, burrowing is even assessed overnight (Mallien et al., 2020). From our experience, mice that might not want to burrow after 2 h or even 6 h might have removed all the substrate from the burrowing tube 24 h later and even started building a nest in the tube, while others are not interacting with the burrowing substrate at all. However, the individual reduction of burrowing

performance is generally considered an indicator of well-being impairments.

Different Stressors Influence Well-Being Differently

We compared two stress-based strategies to elicit depression-associated symptoms in mice. The stress response is mediated by corticosterone, the major glucocorticoid expressed in rodents. Chronic corticosterone has been shown to induce anxiety- and depression-associated symptoms (Strekalova et al., 2004; Gourley et al., 2008; Mitra and Sapolsky, 2008; Zhao et al., 2008; Gourley and Taylor, 2009; Moda-Sava et al., 2019).

In the first approach, the animals are repetitively exposed to the stressful procedure of restraint, in the second approach, the corticosterone is not released physiologically in the mouse after a stressful experience but is instead provided in the drinking water. We used healthy wild-type mice in both strategies and observed their status before, during, and after the stress-based procedures.

The effects of daily *restraint* stress were mild. The BW stagnated during the stress procedure but normalized after the treatment phase in males and females. We also found no change in burrowing performance or TINT. Interestingly, the FCM concentrations did not rise due to the restraint. However, we did see a diminished preference for the saccharin solution in both sexes, from which the mice recovered in the post-stress phase. This indicates a strong emotional burden and a minor effect on physiological parameters. The restraint treatment hardly influences the RELSA. In the RELSA$_{max}$, the restraint stress is the least severe treatment modality in the present study. We deliberately chose a mild restraint paradigm, with only 1 h per day and without an unpredictable or uncontrollable schedule – which is a vital factor in evoking depressive-like states in rodents (Vollmayr and Gass, 2013; Willner, 2017; Strekalova et al., 2022). This might also explain that the preference for saccharin only dropped to ∼75%. Longer restraint phases or unpredictability might increase this effect.

The response to the *corticosterone* treatment was very striking. We observed a rapid weight gain in male and female mice. Although BW alterations concerning animal well-being are typically associated with a loss, we considered this abnormal increase a burden. It is not surprising that the mice showed this increase, given that the treatment is also used to model obesity. However, studies in the literature using 0.1 mg/ml corticosterone in drinking water were only reported in males (Cassano et al., 2012). In addition, we observed untypical behavior in the home cage (limitation of movement, circling) in females after 18 days of treatment. In two cases, the behavioral alteration was so severe that we chose to euthanize the animals as they reached the humane endpoint. We then decided to shorten the procedure from the originally planned 21 days to avoid unnecessary suffering and further losses. The males displayed no such alterations. As soon as the treatment stopped, the BW normalized in both sexes. Of course, as expected, the FCM concentrations increased during the corticosterone treatment but normalized again at the termination of the treatment.

Male *corticosterone* mice showed several other well-being-associated deficits: their burrowing performance decreased during the stress phase. It did not recover during post-stress, their latency to integrate into the nest was increased post-stress, and their preference for saccharin solution was mildly reduced in the post-stress phase. While the *corticosterone*-treated mice in Moda-Sava et al. (2019) dropped from ∼80% preference to 60%, we only found a very mild effect: here, the preference dropped from 88% to only 80%. On the other hand, female corticosterone mice showed no SP or burrowing alterations. This is particularly interesting, as we detected the strong effects in the physiological FCM concentration and BW measures. This leads to the question: Which are the better indicators for the stress burden in this case: the parameters for the psychological or for the physiological burden? In the RELSA score, we used BW, burrowing performance, and FCMs to compare models as they were the most prominent parameters in our analysis. This results in very high RELSA scores in females. Based on this, using males for these specific approaches might be a relevant refinement measure. Especially since the females did not show anhedonic behavior at all.

Based on our data on the *restraint* and *corticosterone ingestion* approaches, we conclude that the *restraint* is a better and more refined model for depression: it evokes the depression-linked symptom of anhedonia with a generally lower burden, especially on the physiological level. In this case, choosing the *restraint* approach over *oral corticosterone* ingestion appears to be a possible refinement.

Comparison of Burden: Stress vs. Genetic Evoked Model for Depression

The other critical observation in this study is comparing the burden of Experiment 1 and Experiment 2: stress-based and genetic alterations. The RELSA score allows the relative comparison of all assessed models. We cannot give an absolute estimate of the burden (mild, moderate, severe), but we can detect which treatment is the most or least detrimental. In the RELSA$_{max}$ (**Figure 5**) the *corticosterone* treatment is the most severe model. All models show elevated RELSA compared to controls. There is no significant difference between the restraint and the GluA1 groups in males, and consequently, their burden is rated as similar. In females, there is a significant difference between the lower RELSA$_{max}$ score of *restraint* stress to GluA1 KO but not to the GluA1 WT. Hence, the GluA1 KO is more severe in females than the daily restraining procedure, while there is no difference between the two procedures in males.

Given that GluA1 KO is likely to be a lifelong burden, while the effects of the stress-based treatments are mostly temporary, we assume that the overall burden is higher in the knockout model. This is not to say upcoming studies should avoid genetic manipulations: GluA1 KO mice show strong psychiatry-associated phenotypes and are an essential tool for gaining more insight into glutamatergic participation in the respective psychiatric disorders (Kilonzo et al., 2022). It is pivotal to have the opportunity to analyze the effects of genes on behavior. However,

it is necessary to consider this burden in the harm-benefit analysis already in the beginning of upcoming projects. Genetic models can and must be justified (Wurbel, 2017; Grimm et al., 2019; Homberg et al., 2021). The researchers should be aware that the choice of a particular genotype might already be a choice against the well-being of the animals. With growing evidence for impaired well-being in genetic lines, researchers have to adjust to the new state of knowledge and accept that models, which formerly were not considered burdened, might be burdened after all. Additionally, the focus should be to alleviate the suffering as much as possible, e.g., taming the animals and avoiding anxiety.

Anxiety in Severity Assessment

Anxiety constitutes a negative affective state and unpleasant feeling that often accompanies the anticipation of events. Of course, such a state impairs the well-being of the animals. Therefore, it is imperative to avoid anxiety in laboratory animals whenever possible. This is pivotal not only for their well-being, but also for the sake of data variability: gentle handling, e.g., cup or tunnel handling, have been shown to reduce anxiety as well as the variation of results at low costs (Hurst and West, 2010; Gouveia and Hurst, 2013, 2017, 2019; Henderson et al., 2020; Sensini et al., 2020). Even though anxiety is a known contributor to laboratory animal distress, we chose to not include it in this study. The assessed parameters were selected to be simple and accessible for everyone. Not every lab has the means and experience to assess anxiety. But also, testing outside the home cage can interfere with the treatment effects (Mallien et al., 2020). Transferring mice from one room to the other, introducing them into aversive environments, e.g., elevated mazes or brightly lit compartments, to evoke an approach-avoidance conflict would most likely cause an additional stress response – which we wanted to avoid in this baseline approach of severity assessment.

Grading Severity Based on Evidence

According to EU directive 63/2010, all animal experiments have to be categorized by the expected degree of pain, suffering, distress, or lasting harm within the project. Concluding from our observations in this study, we cannot make an absolute assessment of severity. However, the animals recovered from the distress in the *restrained* group very quickly and did not show large deflections on the RELSA score. GluA1 KO was similar in males but elevated in females. In comparison with the *corticosterone* treatment, both models were much milder. Therefore, we would consider restraint treatment and GluA1 KO as moderate severity for the animals. Due to the high severity scores and the fact that we had to euthanize two mice in the *corticosterone* treatment group, we think that treatment with 0.1 mg/ml in drinking water for longer than 17 days might also be considered "severe," at least in female mice. That is particularly interesting since other models, e.g., including foot shocks to induce learned helplessness, are considered "severe" according to the assignment criteria in annex VIII. Still, they show only minimal effects on home cage-based well-being parameters (Mallien et al., 2020). Based on these observations, learned helplessness would fall into the "moderate" category, especially since it is also a temporary effect.

In conclusion, we found welfare impairments in all addressed models of depression. This is not surprising as these models are used regularly to inflict symptoms of depression. It is essential to be aware of the burdens of the animals for regulatory and ethical reasons and choose the models wisely. The aim must always be on the benefit side of the harm-benefit balance, but acknowledging the damage and trying to reduce it, e.g., choosing the best model.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by the Regierungspräsidium Karlsruhe.

REFERENCES

AUTHOR CONTRIBUTIONS

AM, NP, CB, DI, RS, RP, ST, and PG: conceptualization and writing – review and editing. ST and AM: formal analysis and visualization. PG: funding acquisition. NP, CB, and AM: investigation. AM and PG: project administration. PG and RS: resources. AM, RP, and PG: supervision. AM, NP, CB, RS, RP, ST, and PG: validation. AM: roles/writing – original draft. All authors contributed to the article and approved the submitted version.

Ang, G., Brown, L. A., Tam, S. K. E., Davies, K. E., Foster, R. G., Harrison, P. J., et al. (2021). Deletion of AMPA receptor GluA1 subunit gene (Gria1) causes circadian rhythm disruption and aberrant responses to environmental cues. *Transl. Psychiatr.* 11:588. doi: 10.1038/s41398-021-01690-3

Austen, J. M., Pickering, C., Sprengel, R., and Sanderson, D. J. (2021). Dissociating Representations of Time and Number in Reinforcement-Rate Learning by Deletion of the GluA1 AMPA Receptor Subunit in Mice. *Psychol. Sci.* 32, 204–217. doi: 10.1177/0956797620960392

Austen, J. M., Sprengel, R., and Sanderson, D. J. (2017). GluA1 AMPAR subunit deletion reduces the hedonic response to sucrose but leaves satiety and conditioned responses intact. *Sci. Rep.* 7:7424. doi: 10.1038/s41598-017-07542-9

Bannerman, D. M., Deacon, R. M., Brady, S., Bruce, A., Sprengel, R., Seeburg, P. H., et al. (2004). A comparison of GluR-A-deficient and wild-type mice on a test battery assessing sensorimotor, affective, and cognitive behaviors. *Behav. Neurosci.* 118, 643–647. doi: 10.1037/0735-7044.118.3.643

Ben Abdallah, N. M., Fuss, J., Trusel, M., Galsworthy, M. J., Bobsin, K., Colacicco, G., et al. (2011). The puzzle box as a simple and efficient behavioral test for exploring impairments of general cognition and executive functions in mouse models of schizophrenia. *Exp. Neurol.* 227, 42–52. doi: 10.1016/j.expneurol.2010.09.008

Birnbacher, D. (2009). "Absolute or relative ethical limits to causing suffering in dealing with laboratory animals," in *The Ethically Acceptable Animal Experiment. Criteria and Boundaries*, ed. D. L. J. Borchers (Paderborn: Brill | mentis).

Bleich, A., Bankstahl, M., Jirkof, P., Prins, J. B., and Tolba, R. H. (2020). Severity Assessment in animal based research. *Lab. Anim.* 54:16.

Bleich, A., and Tolba, R. H. (2017). How can we assess their suffering? German research consortium aims at defining a severity assessment framework for laboratory animals. *Lab. Anim.* 51:667. doi: 10.1177/00236772177733010

Boldt, L., Koska, I., Maarten Van Dijk, R., Talbot, S. R., Miljanovic, N., Palme, R., et al. (2021). Toward evidence-based severity assessment in mouse models with repeated seizures: I. Electrical kindling. *Epilepsy Behav.* 115:107689. doi: 10.1016/j.yebeh.2020.107689

Buchecker, V., Koska, I., Pace, C., Talbot, S. R., Palme, R., Bleich, A., et al. (2022). Toward evidence-based severity assessment in mouse models with repeated seizures: (II.) Impact of surgery and intrahippocampal kainate. *Eur. Surg. Res.* [Epub ahead of print].doi: 10.1159/000522156

Bygrave, A. M., Jahans-Price, T., Wolff, A. R., Sprengel, R., Kullmann, D. M., Bannerman, D. M., et al. (2019). Hippocampal-prefrontal coherence mediates working memory and selective attention at distinct frequency bands and provides a causal link between schizophrenia and its risk gene GRIA1. *Transl. Psychiatr.* 9:142. doi: 10.1038/s41398-019-0471-0

Bygrave, A. M., Masiulis, S., Nicholson, E., Berkemann, M., Barkus, C., Sprengel, R., et al. (2016). Knockout of NMDA-receptors from parvalbumin interneurons sensitizes to schizophrenia-related deficits induced by MK-801. *Transl. Psychiatr.* 6:e778. doi: 10.1038/tp.2016.44

Cassano, A. E., White, J. R., Penraat, K. A., Wilson, C. D., Rasmussen, S., and Karatsoreos, I. N. (2012). Anatomic, hematologic, and biochemical features of C57BL/6NCrl mice maintained on chronic oral corticosterone. *Comp. Med.* 62, 348–360.

Chourbaji, S., Vogt, M. A., Fumagalli, F., Sohr, R., Frasca, A., Brandwein, C., et al. (2008). AMPA receptor subunit 1 (GluR-A) knockout mice model the glutamate hypothesis of depression. *FASEB J.* 22, 3129–3134. doi: 10.1096/fj.08-106450

Deacon, R. M. (2006). Assessing nest building in mice. *Nat. Protoc.* 1, 1117–1119.

Deacon, R. M. (2009). Burrowing: a sensitive behavioural assay, tested in five species of laboratory rodents. *Behav. Brain Res.* 200, 128–133. doi: 10.1016/j.bbr.2009.01.007

Eltokhi, A., Santuy, A., Merchan-Perez, A., and Sprengel, R. (2020). Glutamatergic Dysfunction and Synaptic Ultrastructural Alterations in Schizophrenia and Autism Spectrum Disorder: Evidence from Human and Rodent Studies. *Int. J. Mol. Sci.* 22:59 doi: 10.3390/ijms22010059

Fitzgerald, P. J., Barkus, C., Feyder, M., Wiedholz, L. M., Chen, Y. C., and Karlsson, R. M. (2010). Does gene deletion of AMPA GluA1 phenocopy features of schizoaffective disorder? *Neurobiol. Dis.* 40, 608–621. doi: 10.1016/j.nbd.2010.08.005

Gourley, S. L., and Taylor, J. R. (2009). Recapitulation and reversal of a persistent depression-like syndrome in rodents. *Curr. Protoc. Neurosci.* 9:32. doi: 10.1002/0471142301.ns0932s49

Gourley, S. L., Wu, F. J., Kiraly, D. D., Ploski, J. E., Kedves, A. T., Duman, R. S., et al. (2008). Regionally specific regulation of ERK MAP kinase in a model of antidepressant-sensitive chronic depression. *Biol. Psychiatr.* 63, 353–359. doi: 10.1016/j.biopsych.2007.07.016

Gouveia, K., and Hurst, J. L. (2013). Reducing mouse anxiety during handling: effect of experience with handling tunnels. *PLoS One* 8:e66401. doi: 10.1371/journal.pone.0066401

Gouveia, K., and Hurst, J. L. (2017). Optimising reliability of mouse performance in behavioural testing: the major role of non-aversive handling. *Sci. Rep.* 7:44999. doi: 10.1038/srep44999

Gouveia, K., and Hurst, J. L. (2019). Improving the practicality of using non-aversive handling methods to reduce background stress and anxiety in laboratory mice. *Sci. Rep.* 9:20305. doi: 10.1038/s41598-019-56860-7

Grimm, H., Olsson, I. A. S., and Sandoe, P. (2019). Harm-benefit analysis - what is the added value? A review of alternative strategies for weighing harms and benefits as part of the assessment of animal research. *Lab. Anim.* 53, 17–27. doi: 10.1177/0023677218783004

Hager, C., Keubler, L. M., Biernot, S., Dietrich, J., Buchheister, S., Buettner, M., et al. (2015). Time to Integrate to Nest Test Evaluation in a Mouse DSS-Colitis Model. *PLoS One* 10:e0143824. doi: 10.1371/journal.pone.0143824

Henderson, L. J., Dani, B., Serrano, E. M. N., Smulders, T. V., and Roughan, J. V. (2020). Benefits of tunnel handling persist after repeated restraint, injection and anaesthesia. *Sci. Rep.* 10:14562. doi: 10.1038/s41598-020-71476-y

Homberg, J. R., Adan, R. A. H., Alenina, N., Asiminas, A., Bader, M., and Beckers, T. (2021). The continued need for animals to advance brain research. *Neuron* 109, 2374–2379. doi: 10.1016/j.neuron.2021.07.015

Hurst, J. L., and West, R. S. (2010). Taming anxiety in laboratory mice. *Nat. Methods* 7, 825–826. doi: 10.1038/nmeth.1500

Inta, D., Monyer, H., Sprengel, R., Meyer-Lindenberg, A., and Gass, P. (2010). Mice with genetically altered glutamate receptors as models of schizophrenia: a comprehensive review. *Neurosci. Biobehav. Rev.* 34, 285–294. doi: 10.1016/j.neubiorev.2009.07.010

Jirkof, P., Abdelrahman, A., Bleich, A., Durst, M., Keubler, L., and Potschka, H. (2020). A safe bet? Inter-laboratory variability in behaviour-based severity assessment. *Lab. Anim.* 54, 73–82. doi: 10.1177/0023677721988 1481

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

Kilonzo, K., Strahnen, D., Prex, V., Gems, J., Van Der Veen, B., and Kapanaiah, S. K. T. (2022). Distinct contributions of GluA1-containing AMPA receptors of different hippocampal subfields to salience processing, memory and impulse control. *Transl. Psychiatr.* 12:102. doi: 10.1038/s41398-022-01863-8

Klein, S., Bankstahl, J. P., Loscher, W., and Bankstahl, M. (2015). Sucrose consumption test reveals pharmacoresistant depression-associated behavior in two mouse models of temporal lobe epilepsy. *Exp. Neurol.* 263, 263–271. doi: 10.1016/j.expneurol.2014.09.00

Koska, I., Van Dijk, R. M., Seiffert, I., Di Liberto, V., Moller, C., Palme, R., et al. (2019). Toward evidence-based severity assessment in rat models with repeated seizures: II. Chemical post-status epilepticus model. *Epilepsia* 60, 2114–2127. doi: 10.1111/epi.16330

Kronenberg, G., Balkaya, M., Prinz, V., Gertz, K., Ji, S., and Kirste, I. (2012). Exofocal dopaminergic degeneration as antidepressant target in mouse model of poststroke depression. *Biol. Psychiatr.* 72, 273–281. doi: 10.1016/j.biopsych.2012.02.026

Maksimovic, M., Vekovischeva, O. Y., Aitta-Aho, T., and Korpi, E. R. (2014). Chronic treatment with mood-stabilizers attenuates abnormal hyperlocomotion of GluA1-subunit deficient mice. *PLoS One* 9:e100188. doi: 10.1371/journal.pone.0100188

Mallien, A. S., Hager, C., Palme, R., Talbot, S. R., Vogt, M. A., and Pfeiffer, N. (2020). Systematic analysis of severity in a widely used cognitive depression model for mice. *Lab. Anim.* 54, 40–49. doi: 10.1177/0023677219874831

Mitra, R., and Sapolsky, R. M. (2008). Acute corticosterone treatment is sufficient to induce anxiety and amygdaloid dendritic hypertrophy. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5573–5578. doi: 10.1073/pnas.0705615105

Moda-Sava, R. N., Murdock, M. H., Parekh, P. K., Fetcho, R. N., Huang, B. S., and Huynh, T. N. (2019). Sustained rescue of prefrontal circuit dysfunction by antidepressant-induced spine formation. *Science* 364:eaat8078 doi: 10.1126/science.aat8078

Moller, C., Wolf, F., Van Dijk, R. M., Di Liberto, V., Russmann, V., and Keck, M. (2018). Toward evidence-based severity assessment in rat models with repeated seizures: I. Electrical kindling. *Epilepsia* 59, 765–777. doi: 10.1111/epi.14028

Nestler, E. J., and Hyman, S. E. (2010). Animal models of neuropsychiatric disorders. *Nat. Neurosci.* 13, 1161–1169. doi: 10.1093/ijnp/pyac024

Neumann, I. D., Wegener, G., Homberg, J. R., Cohen, H., Slattery, D. A., and Zohar, J. (2011). Animal models of depression and anxiety: What do they tell us about human condition? *Prog. Neuropsychopharmacol. Biol. Psychiatr.* 35, 1357–1375. doi: 10.1016/j.pnpbp.2010.11.028

Palme, R. (2019). Non-invasive measurement of glucocorticoids: Advances and problems. *Physiol. Behav.* 199, 229–243. doi: 10.1016/j.physbeh.2018.11.021

Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M. T., Baker, M., et al. (2020). The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *BMJ Open Sci.* 4:e100115.

Pittenger, C., Sanacora, G., and Krystal, J. H. (2007). The NMDA receptor as a therapeutic target in major depressive disorder. *CNS Neurol. Disord. Drug Targets* 6, 101–115. doi: 10.2174/187152707780363267

Sanacora, G., Treccani, G., and Popoli, M. (2012). Towards a glutamate hypothesis of depression: an emerging frontier of neuropsychopharmacology for mood disorders. *Neuropharmacology* 62, 63–77. doi: 10.1016/j.neuropharm.2011.07.036

Sanderson, D. J., Good, M. A., Skelton, K., Sprengel, R., Seeburg, P. H., Rawlins, J. N., et al. (2009). Enhanced long-term and impaired short-term spatial memory in GluA1 AMPA receptor subunit knockout mice: evidence for a dual-process memory model. *Learn Mem.* 16, 379–386.

Sanderson, D. J., Lee, A., Sprengel, R., Seeburg, P. H., Harrison, P. J., and Bannerman, D. M. (2017). Altered balance of excitatory and inhibitory learning in a genetically modified mouse model of glutamatergic dysfunction relevant to schizophrenia. *Sci. Rep.* 7:1765.

Seiffert, I., Van dijk, R. M., Koska, I., di liberto, V., Moller, C., Palme, R., et al. (2019). Toward evidence-based severity assessment in rat models with repeated seizures: III. Electrical post-status epilepticus model. *Epilepsia* 60, 1539–1551. doi: 10.1111/epi.16095

Sensini, F., Inta, D., Palme, R., Brandwein, C., Pfeiffer, N., Riva, M. A., et al. (2020). The impact of handling technique and handling frequency on laboratory mouse welfare is sex-specific. *Sci. Rep.* 10:17281. doi: 10.1038/s41598-020-74279-3

Strekalova, T., Liu, Y., Kiselev, D., Khairuddin, S., Chiu, J. L. Y., and Lam, J. (2022). Chronic mild stress paradigm as a rat model of depression: facts, artifacts, and future perspectives. *Psychopharmacology* 239, 663–693. doi: 10.1007/s00213-021-05982-w

Strekalova, T., Spanagel, R., Bartsch, D., Henn, F. A., and Gass, P. (2004). Stress-induced anhedonia in mice is associated with deficits in forced swimming and exploration. *Neuropsychopharmacology* 29, 2007–2017. doi: 10.1038/sj.npp.1300532

Strickland, J. A., Austen, J. M., Sprengel, R., and Sanderson, D. J. (2021). The GluA1 AMPAR subunit is necessary for hedonic responding but not hedonic value in female mice. *Physiol. Behav.* 228:113206. doi: 10.1016/j.physbeh.2020.113206

Talbot, S. R., Struve, B., Wassermann, L., Heider, M., Weegh, N., Knape, T., et al. (2020). One score to rule them all: severity assessment in laboratory mice. *bioRxiv* [Preprint]. doi: 10.1101/2020.06.23.166801

Touma, C., Palme, R., and Sachser, N. (2004). Analyzing corticosterone metabolites in fecal samples of mice: a noninvasive technique to monitor stress hormones. *Horm. Behav.* 45, 10–22. doi: 10.1016/j.yhbeh.2003.07.002

Touma, C., Sachser, N., Mostl, E., and Palme, R. (2003). Effects of sex and time of day on metabolism and excretion of corticosterone in urine and feces of mice. *Gen. Comp. Endocrinol.* 130, 267–278. doi: 10.1016/s0016-6480(02)00620-2

van Dijk, R. M., Koska, I., Bleich, A., Tolba, R., Seiffert, I., Moller, C., et al. (2020). Design of composite measure schemes for comparative severity assessment in animal-based neuroscience research: A case study focussed on rat epilepsy models. *PLoS One* 15:e0230141. doi: 10.1371/journal.pone.0230141

Vollmayr, B., and Gass, P. (2013). Learned helplessness: unique features and translational value of a cognitive depression model. *Cell Tissue Res.* 354, 171–178. doi: 10.1007/s00441-013-1654-2

Weegh, N., Funer, J., Janke, O., Winter, Y., Jung, C., and Struve, B. (2020). Wheel running behaviour in group-housed female mice indicates disturbed wellbeing due to DSS colitis. *Lab. Anim.* 54, 63–72. doi: 10.1177/0023677219879455

Weegh, N., Zentrich, E., Zechner, D., Struve, B., Wassermann, L., and Talbot, S. R. (2021). Voluntary wheel running behaviour as a tool to assess the severity in a mouse pancreatic cancer model. *PLoS One* 16:e0261662. doi: 10.1371/journal.pone.0261662

Wiedholz, L. M., Owens, W. A., Horton, R. E., Feyder, M., Karlsson, R. M., and Hefner, K. (2008). Mice lacking the AMPA GluR1 receptor exhibit striatal hyperdopaminergia and 'schizophrenia-related' behaviors. *Mol. Psychiatr.* 13, 631–640. doi: 10.1038/sj.mp.4002056

Willner, P. (2017). The chronic mild stress (CMS) model of depression: History, evaluation and usage. *Neurobiol. Stress* 6, 78–93. doi: 10.1016/j.ynstr.2016.08.002

Wurbel, H. (2017). More than 3Rs: the importance of scientific validity for harm-benefit analysis of animal research. *Lab. Anim.* 46, 164–166. doi: 10.1038/laban.1220

Zamanillo, D., Sprengel, R., Hvalby, O., Jensen, V., Burnashev, N., Rozov, A., et al. (1999). Importance of AMPA receptors for hippocampal synaptic plasticity but not for spatial learning. *Science* 284, 1805–1811. doi: 10.1126/science.284.5421.1805

Zhao, Y., Ma, R., Shen, J., Su, H., Xing, D., and Du, L. (2008). A mouse model of depression induced by repeated corticosterone injections. *Eur. J. Pharmacol.* 581, 113–120. doi: 10.1016/j.ejphar.2007.12.005