Title:

The behavioral phenotype of early life adversity: a 3-level meta-analysis of preclinical studies

Authors:

V Bonapersona[1,*], J Kentrop[1], CJ Van Lissa[2], R van der Veen[3], M Joëls[1,4], RA Sarabdjitsingh[1]

[1] Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht University, The Netherlands

[2] Department of Methodology and Statistics, Utrecht University, The Netherlands

[3] Centre for Child and Family Studies, Leiden University, The Netherlands.

[4] University Medical Center Groningen, Groningen University, The Netherlands

[*]corresponding author: v.bonapersona-2@umcutrecht.nl

1 ABSTRACT

Background.

Altered cognitive performance has been suggested as an intermediate phenotype mediating the effects of early life adversity (ELA) on later-life development of mental disorders, e.g. depression. Whereas most human studies are limited to correlational conclusions, rodent studies can prospectively investigate how ELA alters cognitive performance in a number of domains. Despite the vast volume of reports, no consensus has yet been reached on the *i)* behavioral domains being affected by ELA and *ii)* the extent of these effects.

Methods.

To test how ELA (here: aberrant maternal care) affects specific behavioral domains, we used a 3-level mixed-effect meta-analysis, a flexible model that accounts for the dependency of observations. We thoroughly explored heterogeneity with MetaForest, a machine-learning data-driven analysis never applied before in preclinical literature. We validated the robustness of our findings with substantial sensitivity analyses and bias assessments.

Results.

Our results, based on >400 independent experiments, yielded >700 comparisons, involving ~8600 animals. Especially in males, ELA promotes memory formation during stressful learning but impairs non-stressful learning. Furthermore, ELA increases anxiety and decreases social behavior. The ELA phenotype was strongest when *i)* combined with other negative experiences ("hits"); *ii)* in rats; *iii)* in ELA models of ~10days duration.

Conclusion.

Prospective and well-controlled animal studies demonstrate that ELA durably and differentially impacts distinct behavioral domains. All data is now easily accessible with MaBapp (https://osf.io/ra947/), which allows researchers to run tailor-made meta-analyses on the topic, thereby revealing the optimal choice of experimental protocols and study power.

2 INTRODUCTION

Early life adversity (ELA) is a consistent risk factor of psychiatric disorders (1, 2), and it is regularly associated with poorer cognitive outcomes later in life (3–5). Indeed, impaired cognitive processing is a prominent feature of psychopathologies (3, 6, 7), e.g. dysregulated contextual memory in post-traumatic stress disorder (8) or social cognition in schizophrenia (9). ELA may therefore alter cognitive development, thereby resulting in behavioral abnormalities that may render individuals more vulnerable to psychiatric disorders (10).

To investigate exactly how ELA affects cognitive processing, rodent models are a valuable resource: they complement human studies by in-depth and thorough investigations of otherwise hard-to-study mechanisms. In animal experiments, genetic and environmental influences can be more precisely controlled and experimentally varied than in humans (11). Furthermore, prospective designs are more feasible. For example, rodent studies have disentangled the different components of mother-pup interaction, a critical factor of early development across mammalian species (12, 13). This has helped uncover links between disturbed maternal care and disturbed emotional and cognitive functioning later in life, implicating the stress system (14) and "hidden regulators" (15).

Rodent studies have also highlighted paradoxical ELA effects on cognitive abilities. For instance, Benetti *et al.* (16) reported that rats with a history of ELA had impaired memory in the object recognition task. Conversely, Champagne *et al.* (17) reported that ELA mice display increased memory in a fear conditioning paradigm. Both tests have historically been used as memory tasks, albeit in a non-stressful and stressful context respectively. Possibly, the equivocal results are due to different underlying biological mechanisms (e.g. learning in stressful vs non-stressful situations) or pertain to the divergent methodology used (e.g. type of test or ELA model, species, experimenters, labs). A few studies have investigated the latter by testing the same animals in different memory tasks (18–21). Although these studies favor the former explanation, the limited amount of animals used (22)– alongside the heterogeneous methodology– prohibits firm conclusions.

To address this conundrum, we here carried out a large-scale 3-level meta-analysis of all peer-reviewed preclinical literature on the subject, and tested the hypothesis that ELA (particularly aberrant maternal care) differentially affects specific behavioral domains in adulthood. We focused on memory formation after stressful or non-stressful learning, anxiety and social behavior, given their relevance for psychopathologies. We addressed (potential) sex-differences by investigating males and females separately. Furthermore, we tested whether the presence of multiple hits (23) (e.g. other negative life

experiences) amplified ELA effects. Finally, we applied a novel, machine-learning based analysis (MetaForest(24)) to identify the most important moderators of ELA effects on behavior.

Based on this comprehensive analysis, we evaluate the translational potential of ELA rodent models. With the aid of a specially developed web-based tool MaBapp (*Meta-Analysis of Behavior application*) (https://osf.io/ra947/), interested researchers can perform their own meta-analysis and retrieve valuable *ad hoc* information for experimental design and power calculations.

3   METHODS

We adhered to SYRCLE's guidelines (25)(26), and to the PRISMA (27) reporting checklist. To ease reading of the methodology, definitions of technical terms are provided in Supplemental Methods (S1.1). A summary of the general approach can be found in Figure 1.

3.1   SEARCH STRATEGY

The electronic databases PubMed and Web of Science (Medline) were used to conduct a comprehensive literature search on *the effects of ELA on behavior* on December 6th 2017. The search string was constructed with the terms "behavioral tests", "ELA" and "rodents" (S1.2).

Prior to the beginning of the study, four experts were consulted. After elaborate discussions they agreed upon *i)* the selection of tests and related outcomes (S1.3), *ii)* their classification into behavioral domains (S1.3) and *iii)* the definition of multiple hits (S1.4). The results of each individual test, independent of categorization, are available for consultation on MaBapp (Section 5.1). Studies' titles and abstracts were screened independently by two researchers (VB & JK) and selected if the inclusion criteria were met (S1.5). Studies' inclusion was performed blinded to the studies' results. In case of doubt, the full text was inspected. Any disagreement was resolved by greater scrutiny and discussion.

To limit subjectivity in the data gathering and entry process, data from eligible studies were extracted in a standardized dataset alongside its explanatory codebook (https://osf.io/ra947/).

For each individual comparison, we calculated *Hedge's* G (28), a standardized mean difference with a correction for small samples (29). S1.6 details the extraction of statistical information as well as handling of missing values. We estimated the summary statistics of data presented only graphically with *Ruler for Windows* (https://a-ruler-for-windows.en.softonic.com/), of which we previously validated the accuracy (30). If the data was not reported in any format (or other crucial information was missing e.g. sex), we contacted two authors per manuscript published after 2008 (response rate 52.6%). If no answer was received within two months and after a reminder, the authors were considered not reachable, and the comparison was excluded.
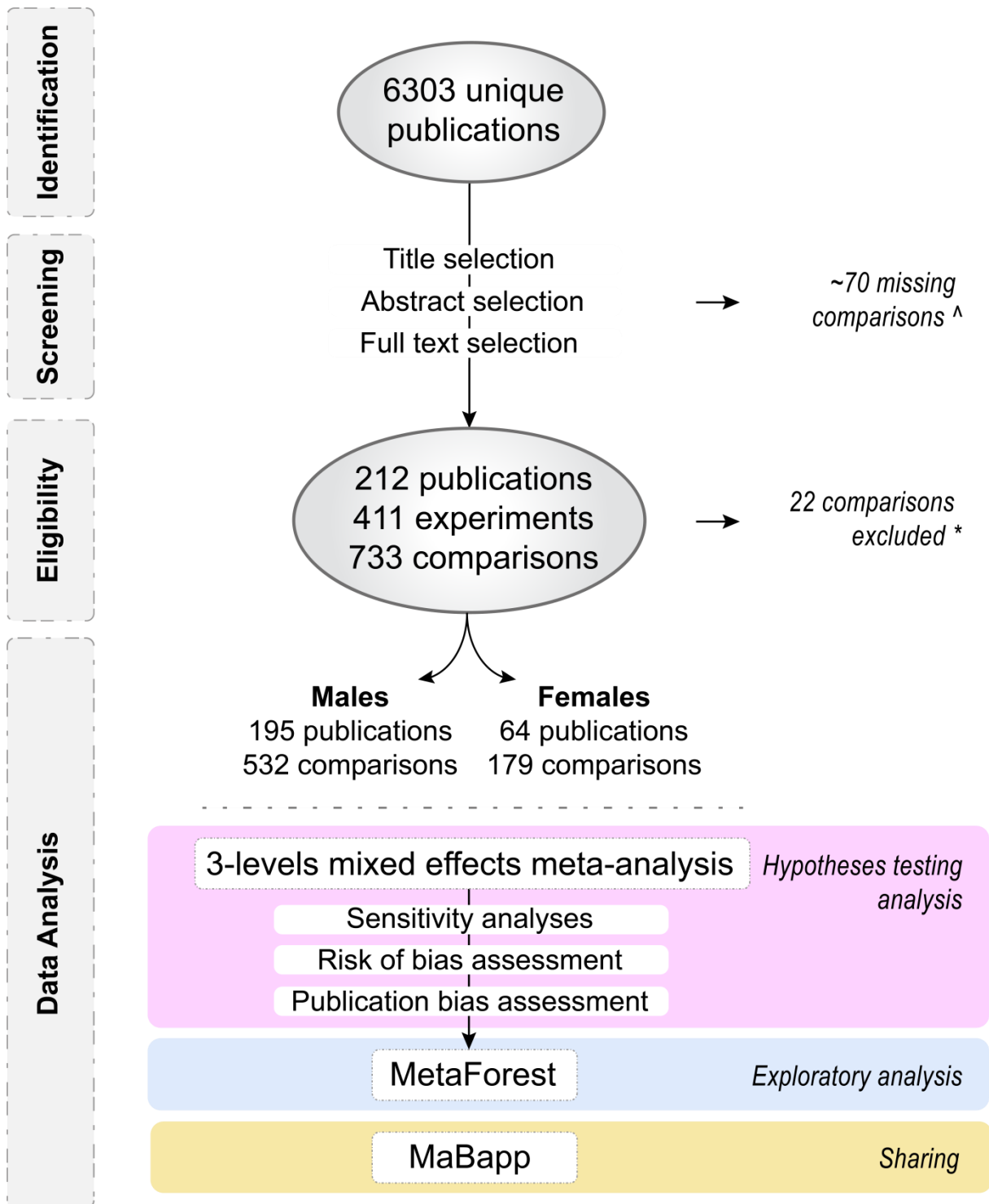
Figure 1 | Flow chart of study selection and analysis. ^ = estimation of missing comparisons (S2.1); * = comparisons excluded from the meta-analysis due to controversial behavioral domain categorization (S2.2)

3.2 META-ANALYSIS: RESEARCH QUESTIONS AND STATISTICAL APPROACH

To avoid possible biases, the experimenter (VB) was blinded to the ELA effects while coding the analysis. This was achieved by randomly multiplying half of the effect sizes by -1.

3.2.1   Hypothesis-testing

We built a 3-level mixed effect meta-analysis with restricted maximum likelihood estimation. In our experimental design, the 3 levels correspond to variance of effect size between 1) animals, 2) outcomes and 3) experiments. This approach accounts for the violation of the assumption of independency when the data is collected from the same animals (30–32), thereby improving the robustness of the conclusions drawn. We included "domains" and "hits" as moderators in order to address the following two research question: 1) *what are the effects of ELA on each behavioral domain?*; 2) *are the effects enhanced if the animals experienced multiple hits?*. Since both questions were answered with the same model, effect sizes were estimated only once.

We ensured that all behavioral measurements were in the same theoretical direction by multiplying – whenever necessary – the effect sizes by -1 (29)(S1.3). Although this was essential for the model estimation, we here report effect sizes in a more interpretable manner: an increase in Hedge's G signifies an enhancement of the behavioral domain under study (e.g. more anxiety, better memory).

We conducted several sensitivity analyses (S1.7) to assess the robustness and consistency of our conclusions. We examined whether the quality of the studies affected the estimation of the results by dissecting the influence of reporting bias, blinding, randomization and study power. Furthermore, we thoroughly investigated influential and outlying cases (33) according to multiple definitions (S1.7).

To compensate for methodological limitations, we tested the presence of publication bias with various qualitative/quantitative methods (S1.8), and quantified its influence with fail-safe N (34) and trim-and-fill analyses (35)(S1.8).

Risk of bias was evaluated with SYRCLE's assessment tool (36), where we distinguished between study-level and outcome-level biases (27). Lack of reporting of experimental details was scored as an *unclear* risk of bias.

Heterogeneity was assessed with Cochrane Q-test (32) and I^2, which was estimated at each of the 3-levels of the model to determine how much variance could be attributed to differences within (level 2) or between experiments (level 3) (37). Estimates of explained variance can be positively biased when based on the data used to estimate the model (38). For this reason, we used 10-fold cross-validation to obtain an estimate of how much variance our model might explain in new data. This cross-validated estimate of R^2 (R_{cv}^2) is robust to overfitting and provides evidence for the results' generalizability.

P-values were corrected with Bonferroni for family-wise error rate (each research question considered as a separate family of tests) to limit capitalization on chance. Since we expected the amplitude of effect sizes to differ between sexes (39, 40), we considered males and females as two separate datasets.

3.2.2   Exploratory analysis

We used MetaForest (24), an novel exploratory approach to identify the most important moderators of the ELA effects on behavioral domains. This innovative, data-driven technique adapts random forests (a machine learning algorithm) for meta-analysis, by means of bootstrap sampling. MetaForest ranks moderators based on their influence on the effect size.

Preclinical experiments often adopt diverse protocols. Although this can be an advantage (41), in a meta-analysis it induces substantial heterogeneity. Therefore, we classified the published experimental protocols in >30 standardized variables with the intent to identify potential methodological sources of heterogeneity. Based on theoretical importance, we selected 26 of these moderators for inclusion in the MetaForest analysis. We used 10-fold cross-validation (S1.9) to determine the optimal tuning parameters that minimized RMSE: uniform weighting, 4 candidate moderators at each split, and a minimum node size of 2. The marginal bivariate relationship of each moderator with effect size was averaged over the values of all other moderators (S1.9). Residual heterogeneity was estimated with tau^2 (S1.9).

Lastly, we created MaBapp (https://osf.io/ra947/) for anyone to perform their own meta-analysis on the topic by selecting their favorite characteristics (Section 5.1).

Analyses were conducted in R (version 3.5.1) (42), using the following packages: 1) *metafor* (28) for conducting the analysis, 2) *metaforest* (43) for data exploration, 3) *shiny* (44) to create MaBapp, and 4) *dplyr* (45) for general data handling. For further specifications about the analysis, the R script is made available (https://osf.io/ra947/).

4 RESULTS

4.1 STUDIES SELECTION AND CHARACTERISTICS

In total ~8600 animals (age$_{weeks}$ median[IQR]=12[4]; proportion rats=68%) were included in the analysis, 77.7% of which were males. Anxiety was the behavioral domain most investigated (48.8%), elevated plus maze the most popular test (14.3%), and maternal separation the ELA paradigm most often used (48.9%). For additional descriptive information on study characteristics, see S2.3.

Although no publication reported on all SYRCLE's potential bias items, 41 publications (19.3%) were blinded as well as randomized, and overall we estimated a risk of bias of 3[1] (median[IQR]) on a 10 points scale (S2.4). Lastly, at a systematic review level (S2.5), 68.5% of comparisons were either not-significant (n$_{comp}$=386) or the result could not be directly interpreted from the information provided (n$_{comp}$=117).

4.2 ELA EFFECTS ARE PRONOUNCED IN MALES AND WITH "MULTIPLE HITS"

The effect sizes included ranged between -6.4 and 6.1 (mean[SD]=.29[1.06]), with 95% of comparisons between -2 and 2. Sample size ranged between 6 and 59 animals (mean[SD]=22[7.8]), and differed <20% between control and ELA groups in 90% of the cases (estimation).

When qualitatively comparing sexes, the effects of ELA were more evident in males than in females. **Male** rodents with a history of ELA displayed increased anxiety (HedgesG[95%CI]=.278[.165,.39], z=4.819, p<.000), improved memory after stressful learning (HedgesG[95%CI]=.283[.141,.425], z=3.9, p<.000), impaired memory after non-stressful learning (HedgesG[95%CI]=-.594[-.792,-.395], z=-5.86, p<.000) and decreased social behavior (HedgesG[95%CI]=-.614[-.88,-.348], z=-4.521, p<.000, Figure **Error! Reference source not found.**A, S2.6). We were unable to confirm any effect of ELA on behavior in **females**, although directionality was generally comparable in both sexes (**Error! Reference source not found.**B, S2.7).

Overall, the presence of multiple hits intensified the effects of ELA in **males** (HedgesG[95%CI]=.222[.018,.426], z=2.131, p=.033) yet marginally in **females** (HedgesG[95%CI]=.297[-.003,.596], z=1.939, p=.052). Although these enhancing effects were not significant at a single-domain level (posthoc analysis) (Figure 2C-D, S2.6/S2.7), memory after non-stressful learning was the most impacted domain in **males** (difference in Hedge's G=.435, z=2.156, p=.124) as well as in **females** (difference in Hedge's G=.565, z=2.234, p=.102).
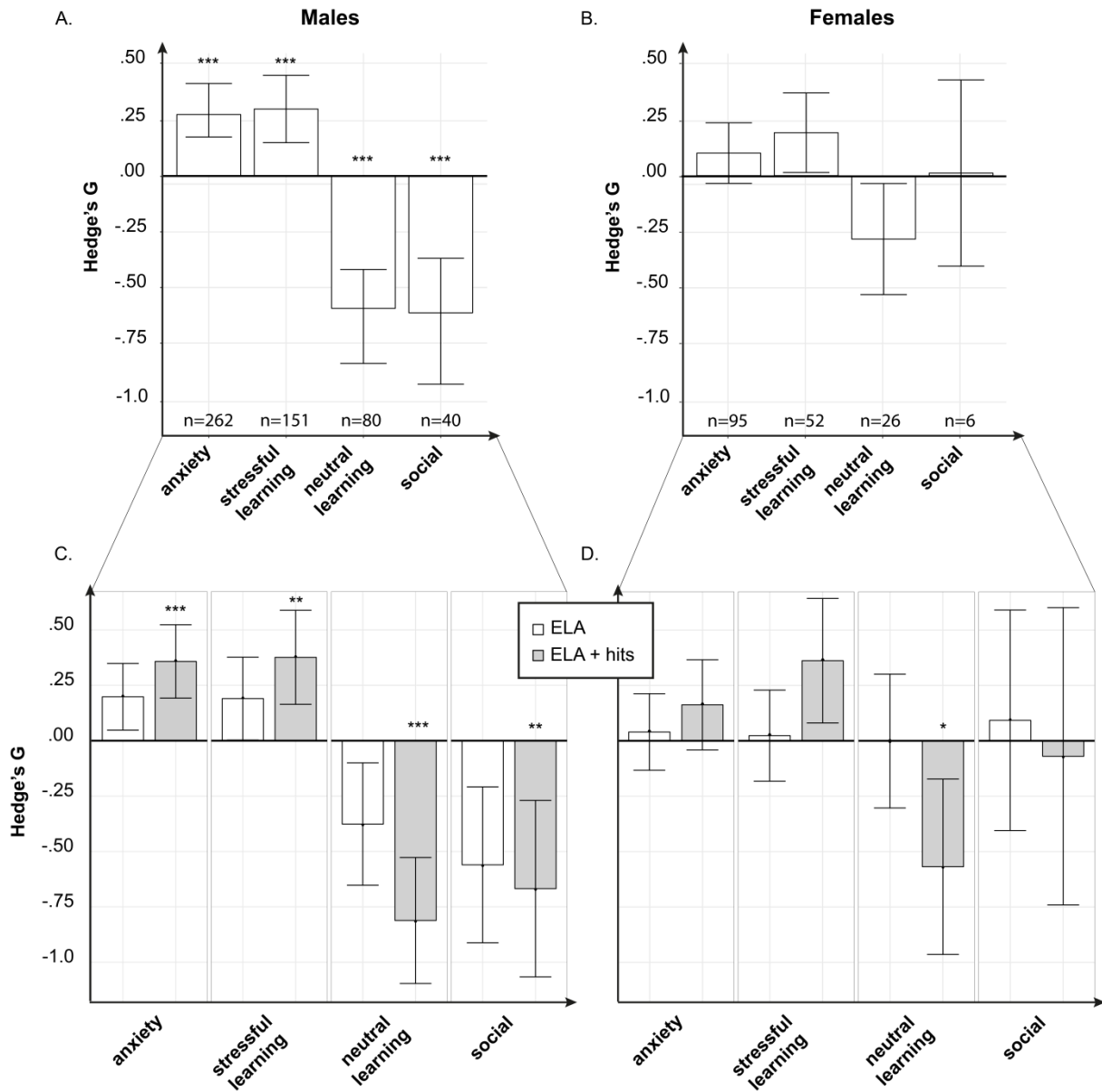
Figure 2 | Effects of ELA on behavioral domains in males (**A**) and females (**B**), and the role of multiple hits (in addition to ELA, grey bars) compared to only ELA (white bar) in mediating these effects (males: **C**, females: **D**). Each bar represents the size of the effect (*Hedge's G*, standardized mean difference) of the ELA manipulation when comparing a control and an experimental group. * = p<.05, ** = p<.01, *** = p<.001

4.2.1 Sensitivity analyses and publication bias.

Qualitative evaluation of funnel plot asymmetry suggested the presence of publication bias, which was confirmed by Egger's regression and Begg's test (S2.8). Nonetheless, fail-safe N as well as trim-and-fill analyses confirmed that – albeit present – publication bias is unlikely to distort the interpretation of the results (S2.8). Furthermore, the robustness of the male and female models was confirmed by several sensitivity analyses (S2.9).

4.3 EXPLORATION OF MODERATORS

Although the models of the hypotheses-testing analysis described a significant proportion of variance ($R_{cv}^2{}_{males}$=.026, $R_{cv}^2{}_{females}$=.03), substantial heterogeneity was recorded in both models (**males**: Q(524)=1763.118, p<.000; **females**: Q(171)=326.93, p<.000, S2.10). This was not surprising due to the diversity of publications included in the meta-analysis.

To investigate the source of the heterogeneity, we used MetaForest, a new statistical technique that ranks moderators (Figure 3A) based on their predictive value. These can roughly be divided in 4 groups, describing: *i)* characteristics of the animals (e.g. origin of the breeding animals (Figure 3B) and species investigated (Figure 3C)), *ii)* ELA model used (e.g. type of model and duration of ELA (Figure 3D-E)), *iii)* outcome measures (e.g. domain and test used), and iv) potential bias (e.g. blinding and randomization). MetaForest confirmed that the selected moderators account for a substantial portion of the variance (R_{cv}^2[SD]=.12[.09]).

Offspring of dams purchased pregnant had larger effect sizes than offspring bred in the own facility (Figure 3B). Rats had overall larger effect sizes than mice (Figure 3C). Concerning ELA models (Figure 3D), selecting the extremes of natural variation (licking-and-grooming model) yielded the strongest phenotype. Lastly, effect sizes appeared to be maximal with a 10 days' ELA duration (Figure 3E).
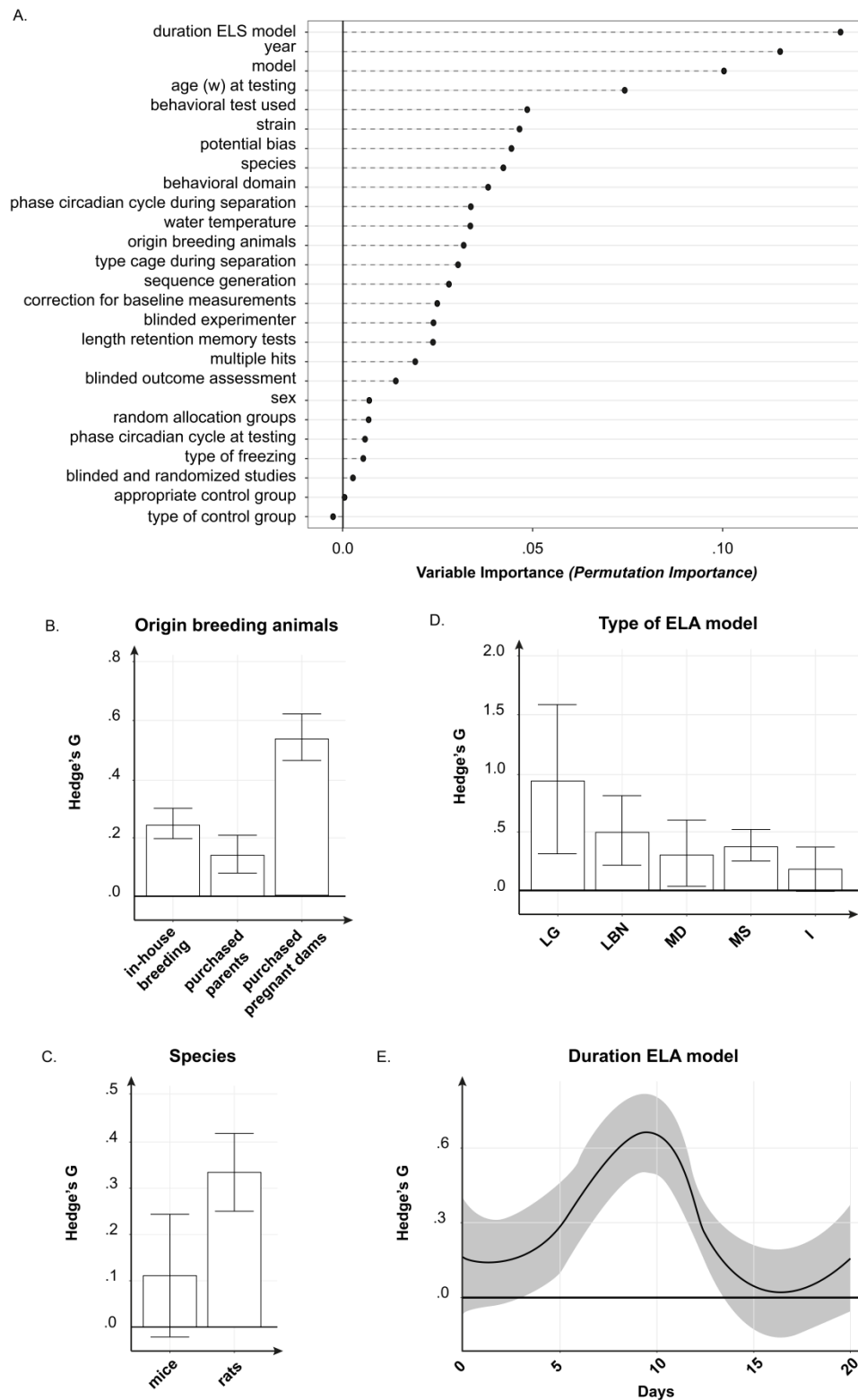
11

Figure 3 | Exploratory MetaForest analysis. (**A**) Rank moderators' importance. Variable/permutation importance is a measure of how strongly each moderator explains differences in effect size, capturing (non-)linear relationships as well as higher order interactions. For information about MetaForest's partial dependence plots, see S2.11. Effect sizes distinguished by origin of the breeding animals (**B**), species (**C**), type of ELA model (**D**) and duration of ELA (**E**). Results are expressed as Hedge's G[95%CI]. The usefulness of this exploration can be best appreciated with the aid of MaBapp. For example, the overall estimate of the effects of ELA on anxiety is Hedge's G=.24. However, if we select only the LBN model, the effect size rises to .37. If we combine LBN and rats, the effect size further rises to .60. If we then select only elevated plus maze as respectively behavioral test, the effect size rises to .81. LG = licking-and-grooming, LBN = limited bedding and nesting, MD = maternal deprivation, MS = maternal separation, I = isolation.

5 DISCUSSION

In this study, we substantiate that adversities early in life profoundly and lastingly change rodent behavior. Due to low power (22) and heterogeneous methodologies, results at a single-study level are often inconclusive and difficult to interpret. Here, by adopting a meta-analytic approach, we provide extensive evidence that ELA has differential effects on memory: it enhances memory if learning occurs in a stressful situation, but it hampers learning under non-stressful circumstances. Furthermore, ELA increases anxiety and decreases social behavior, particularly in males. In line with the multiple-hits hypotheses (23, 46), the effects are amplified if the animals experience other stressful life events (e.g. prenatal stress due to transport of pregnant females). These results are independent of the type of ELA or behavioral test used, and are remarkably similar to what was reported at a correlational level in humans (47, 48). Altogether, our results point to a clear and robust phenotype of ELA in four behavioral domains and complement the human literature by supporting a causative role of ELA in altering behavior, which may predispose individuals to precipitate symptoms of psychiatric disorders.

5.1 METHODOLOGICAL CONSIDERATIONS

The lack of sufficient power to detect experimental effects is an emerging issue in preclinical literature (22, 30) that seriously hampers research interpretation (22). As a consequence, results from single-studies are useful for hypotheses generation but do require replication. Indeed, in our study the majority of comparisons (68.5%) was not-significant at a systematic review level, but the effects were significant when analyzed meta-analytically. Although still sparsely applied to preclinical studies, statistical tools such as meta-analyses can therefore be very useful to substantiate conclusions from animal data and translate them more reliably to patients (49).

In this project, we intertwine this concept with state-of the-art statistical methodology, adopting an approach never used in preclinical studies. Firstly, our meta-analysis was built with a 3-level model (32), which allows for a more robust estimation of the effects by accounting for the dependency of same-animal's data (30, 50). Secondly, a leading strength of preclinical meta-analyses is the systematic exploration of heterogeneity (49). Instead of the standard subgroup/meta-regression approach, we opted for an exploratory analysis using MetaForest (24), a newly developed technique that ranks moderators' importance by adapting the machine learning algorithm *random forests* to summary-statistics' data. A major strength of MetaForest is its robustness to overfitting, and its ability to accommodate non-linear effects, as shown by the impact of ELA duration on effect sizes (24).

Thirdly, we extensively coded potential (biological and experimental) moderators. Although possibly relevant moderators were not included due to insufficient reporting (e.g. temperature during

separation (51), cross-fostering (52), culling (53)), this dataset treasures relevant information for future experimental designs. To facilitate others to exploit this dataset, we created **MaBapp** (https://osf.io/ra947/), a web-based app with a user-friendly interface through which anyone can perform his/her own meta-analysis on the topic of ELA and behavioral domains. Within the app, a wide variety of features can be selected, such as ELA models and their components (e.g. type, timing, predictability), behavioral tests used, age and sex of the animals, etc. Based on the characteristics indicated, the app reports forest, funnel and cumulative plots, as well as a list of relevant publications. The app is a useful resource, which can be used to i) comprehensively retrieve relevant publications, ii) explore the literature at an individual researcher's needs' level, iii) define new hypotheses, iv) evaluate publication bias and replicability of findings, and v) estimate realistic effect sizes on which to ground future research.

The validity of our conclusions is not limited to the robustness of the models used but grounded on the vast primary evidence included (>200 publications). As a consequence, accidental findings have little weight. Although the methods and approach we adopt are rigorous and reasonably conservative, the quality of the conclusions critically depends on the quality of the studies and data included. From our qualitative bias assessment, the risk for potential bias was lower than previously reported in Neuroscience (30, 54, 55); yet, only ~20% of studies stated being blinded as well as randomized. Furthermore, any meta-analytic dataset is burdened with missing data, due to publication bias or to the preferred investigation of certain factors over others (56). Our models did display evidence of publication bias, yet they were robust to several corrections and sensitivity analyses. Although we cannot fully exclude that the above-mentioned limitations may affect the outcome, it is unlikely that the conclusions drawn would be substantially impacted. Nevertheless, we have attempted to address these methodological issues as comprehensively as possible in our analysis.

5.2 CONSIDERATIONS ON ELA MODELS

The behavioral changes we report are presumably a convergent phenotype of distinct, model-dependent, underlying biological mechanisms. An organism's development is not linear nor simultaneous for every component, but it occurs in critical periods (57). For example, postnatal day (P)2-P5 is a sensitive period for the maturation of the adrenal glands (58), P9 for prepulse inhibition (59), and ~P10 for adrenal responsiveness (60). Furthermore, higher cognitive functions develop as multistage processes of sequential nature (57). Accordingly, ELA may particularly disrupt the development of competences whose critical period is active during the time of stress, thereby heightening the variability of the ELA phenotype.

Evidence supporting these notions derives from studies using a single 24h maternal deprivation paradigm, which show a persistent yet paradoxical hypo- and hyper- responsiveness of juvenile ACTH if deprivation occurred at P3 or P11 respectively (61). Thus, while meta-analyses may serve to discern patterns among vast amounts of studies, exploratory studies experimentally dissecting components of ELA in rodents remain indispensable for addressing the underlying mechanisms of action of ELA to the brain (for example: (62, 63)).

5.2.1   Suggestions for future ELA research

Given that the criteria for construct and face validity of ELA models have been met (64), our results provide a practical framework where researchers can anticipate the ELA effect on cognitive outcomes and/or build their own ELA model accordingly. Our exploratory analysis gives insights in the suitability of the models and tests to choose, depending on the question.

Based on this analysis, we tentatively conclude that *i)* rats seem overall more sensitive to ELA-induced changes than mice. Moreover, *ii)* elements such as transporting pregnant dams appear to amplify the effects of ELA. Such stressful life events may have substantial impact on the system, in line with the multiple-hit theory (23). As evident from Figure 3, *iii)* a duration of ~10 days ELA produced the most robust phenotype. Finally, *iv)* the limited bedding and nesting (LBN) model produced the largest effect sizes when compared to separation/deprivation models. Given this reliability, in combination with the feasibility and translational validity, LBN seems an influential paradigm to investigate the mechanisms of chronic stress early in life (39, 65).

To reduce variability and improve comparability across studies, ELA should be preferably applied with consistent protocols (S1.5), unless manipulation of particular aspect(s) of the model is under investigation. Obviously, the importance of individual variation is a factor that should not be overlooked. In our analysis, the paradigm of licking-and-grooming – which is not experimentally induced but based on natural variation in maternal care – consistently evoked the largest effect sizes, although these were based on fewer publications than the other models.

5.3   TRANSLATIONAL POTENTIAL

ELA is one of the most consistent environmental risk factors for the development of psychopathology (2). Although the effects of ELA on the brain can be adaptive, they may evolve into dysfunctional elements in genetically predisposed individuals (2). Behavioral performance in specific cognitive domains seems to be a relevant intermediate phenotype (10), as it may mediate the effects of ELA on

psychopathology. For example, in post-traumatic stress disorder, enhanced memory of stressful events becomes pathological after a later-life trauma (8).

In humans, the concept of ELA is extremely varied. Even when considering solely maltreatment, this can be characterized by repeated or sustained episodes of various forms of neglect and abuse (66). Furthermore, the environmental variation is intertwined with socio-economic status, complex relations (e.g. family, neighborhoods, peers, school), and intergenerational transmissions (66). Rodent paradigms do not capture the complexity of human ELA, but they can model specific *aspects* of the human variability in a well-controlled setting. For example, LBN is based on the erratic and unpredictability of maternal care (39, 65, 67) that has been established as a hallmark in childhood abuse situations (68). Similarly, cognitive performance (e.g. memory after stressful learning) can be modelled in rodents, albeit with clear restraints: the tasks are obviously different, should be interpreted in relation to the animal's normal behavior, and cannot investigate a range of outcomes such as verbal abilities, critical for social interaction and psychopathology (69), also in relation to ELA (70).

Explaining how ELA increases psychopathology risk requires the understanding of its complex interplay with other susceptibility/resilience factors, such as genetic background and later life stressors (71). This mechanistic investigation is difficult to achieve in humans, where limited material, difficulty of prospective and longitudinal designs, complexity and lack of control over the environment and genetic variation hamper causal inferences of ELA to later life cognitive performance. To this end, animal studies can be of considerable added value (39).

An interesting issue in evaluating the translational potential of ELA rodent models is sex differences. In our analysis, males showed larger effect sizes (albeit in the same direction) than females to the effects of ELA on all outcomes, thereby confirming previous preclinical literature (72). Conversely, in clinical populations, females appear more sensitive to childhood trauma as well as to the development of stress-related psychopathologies (39), although sex differences depend on the type of disorder (73). A plausible biological explanation for this discrepancy is the developmental timing during which stress occurs. Although humans and rodents are altricial species, the brain of newborn rats corresponds roughly to 23/24-week old human fetuses (74). Interestingly, the sensitivity to adversities in the last trimester of gestation in humans has been suggested to affect males more than females (75). Experimentally manipulating the timing of ELA exposure may further elucidate 'female' stress-sensitive periods. It therefore remains to be established whether the effects of ELA on cognitive domains are truly different between sexes. Our analyses suggest that the effects may not be sexually dysmorphic in nature but may result from the experimental designs used. For example, ELA models

and behavioral tests were originally developed for males: maternal care shows clear sex-specific differences (76, 77), and females perform poorly in behavior tests such as object recognition and object-in-location (39, 40). Consequently these paradigms may not be sensitive enough for a female's phenotype. Indeed, the recorded effects were in the same direction across sexes, and MetaForest attributed to sex a relatively modest variable importance. Our results showcase the necessity to study sex as a biological variable (75, 78), which requires the development of tests and models that are female-specific. This step is required for a more meaningful comparison between rodent and humans, and a delineation of the underlying sex-dependent mechanisms of ELA.

Despite these drawbacks, our meta-analysis confirms and importantly extends standing hypotheses on ELA based on exploratory studies. To aid future investigations in this field, we provide a online tool to evaluate existing literature and direct the experimental design of new studies.

6 ACKNOWLEDGEMENTS

7 BIBLIOGRAPHY

1. Kessler RC, McLaughlin KA, Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, *et al.* (2010): Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *Br J Psychiatry*. 197: 378–85.

2. Teicher MH, Samson JA, Anderson CM, Ohashi K (2016): The effects of childhood maltreatment on brain structure, function and connectivity. *Nat Rev Neurosci*. 17: 652–666.

3. Masson M, East-Richard C, Cellard C (2016): A meta-analysis on the impact of psychiatric disorders and maltreatment on cognition. *Neuropsychology*. 30: 143–156.

4. Nelson CA, Zeanah CH, Fox NA, Marshall PJ, Smyke AT, Guthrie D (2007): Cognitive Recovery in Socially Deprived Young Children: The Bucharest Early Intervention Project. *Science (80-)*. 318: 1937–1940.

5. Vargas T, Lam PH, Azis M, Osborne KJ, Lieberman A, Mittal VA (2018): Childhood Trauma and Neurocognition in Adults With Psychotic Disorders: A Systematic Review and Meta-analysis. *Schizophr Bull*. 1–14.

6. Millan MJ, Agid Y, Brüne M, Bullmore ET, Carter CS, Clayton NS, *et al.* (2012): Cognitive dysfunction in psychiatric disorders: Characteristics, causes and the quest for improved therapy. *Nat Rev Drug Discov*. 11: 141–168.

7. Monfils MH, Holmes EA (2018): Memory boundaries: opening a window inspired by reconsolidation

to treat anxiety, trauma-related, and addiction disorders. *The Lancet Psychiatry*. 0366. doi: 10.1016/S2215-0366(18)30270-0.

8. Liberzon I, Abelson JL (2016): Context processing and the neurobiology of post-traumatic stress disorder. *Neuron*. 92: 14–30.

9. Green MF, Horan WP, Lee J (2015): Social cognition in schizophrenia. *Nat Rev Neurosci*. 16: 620–631.

10. Ammerman RT, Cassisi JE, Hersen M, Van Hasselt VB (1986): Consequences of physical abuse and neglect in children. *Clin Psychol Rev*. 6: 291–310.

11. Knop J, Joëls M, van der Veen R (2017): The added value of rodent models in studying parental influence on offspring development: opportunities, limitations and future perspectives. *Curr Opin Psychol*. 15: 174–181.

12. Bowlby J (1951): Maternal care and mental health. *World Heal Organ Monogr Ser*. .

13. Harlow H, Harlow M (1965): The affectional systems. In: Schrier A, Harlow H, Stollnitz F, editors. *Behav Nonhum primates*, 1st ed. New York: Academic Press, pp 287–334.

14. Levine S (2005): Developmental determinants of sensitivity and resistance to stress. *Psychoneuroendocrinology*. 30: 939–946.

15. Hofer MA (1978): Hidden Regulatory Processes in Early Social Relationships. In: Bateson P, Klopfer P, editors. *Perspect Ethol Soc Behav*. New York: Penum, pp 135–201.

16. Benetti F, Mello PB, Bonini JS, Monteiro S, Cammarota M, Izquierdo I (2009): Early postnatal maternal deprivation in rats induces memory deficits in adult life that can be reversed by donepezil and galantamine. *Int J Dev Neurosci*. 27: 59–64.

17. Champagne DL, Bagot RC, van Hasselt F, Ramakers G, Meaney MJ, de Kloet ER, *et al.* (2008): Maternal Care and Hippocampal Plasticity: Evidence for Experience-Dependent Structural Plasticity, Altered Synaptic Functioning, and Differential Responsiveness to Glucocorticoids and Stress. *J Neurosci*. 28: 6037–6045.

18. Kanatsou S, Karst H, Kortesidou D, van den Akker RA, den Blaauwen J, Harris AP, *et al.* (2017): Overexpression of Mineralocorticoid Receptors in the Mouse Forebrain Partly Alleviates the Effects of Chronic Early Life Stress on Spatial Memory, Neurogenesis and Synaptic Function in the Dentate Gyrus. *Front Cell Neurosci*. 11: 1–13.

19. Ivy AS, Rex CS, Chen Y, Dube C, Maras PM, Grigoriadis DE, *et al.* (2010): Hippocampal Dysfunction

and Cognitive Impairments Provoked by Chronic Early-Life Stress Involve Excessive Activation of CRH Receptors. *J Neurosci*. 30: 13005–13015.

20. Bredy TW, Humpartzoomian RA, Cain DP, Meaney MJ (2003): Partial reversal of the effect of maternal care on cognitive function through environmental enrichment. *Neuroscience*. 118: 571–576.

21. Mello PB, Benetti F, Cammarota M, Izquierdo I (2009): Physical exercise can reverse the deficit in fear memory induced by maternal deprivation. *Neurobiol Learn Mem*. 92: 364–369.

22. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013): Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 14: 365–376.

23. Daskalakis NP, Bagot RC, Parker KJ, Vinkers CH, de Kloet ER (2013): The three-hit concept of vulnerability and resilience: Toward understanding adaptation to early-life adversity outcome. *Psychoneuroendocrinology*. 38: 1858–1873.

24. van Lissa CJ (2018): MetaForest: Exploring heterogeneity in meta-analysis using random forests. *PsyArXiv*. . doi: 10.31234/osf.io/myg6s.

25. De Vries RBM, Hooijmans CR, Langendam MW, Leenaars M, Ritskes-Hoitinga M, Wever KE (2015): A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclin Med*. 1: 1–9.

26. Leenaars M, Hooijmans CR, van Veggel N, ter Riet G, Leeflang M, Hooft L, *et al.* (2012): A step-by-step guide to systematically identify all relevant animal studies. *Lab Anim*. 46: 24–31.

27. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, *et al.* (2009): Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*. 6. doi: 10.1371/journal.pmed.1000097.

28. Viechtbauer W (2010): Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw*. 36: 1–48.

29. Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, *et al.* (2014): Meta-analysis of data from animal studies: A practical guide. *J Neurosci Methods*. 221: 92–102.

30. Bonapersona V, Joëls M, Sarabdjitsingh RA (2018): Effects of early life stress on biochemical indicators of the dopaminergic system: a 3 level meta-analysis of rodent studies. *Neurosci Biobehav Rev*. 95: 1–16.

31. Aarts E, Verhage M, Veenvliet J V, Dolan C V, van der Sluis S (2014): A solution to dependency: using multilevel analysis to accommodate nested data. *Nat Neurosci*. 17: 491–496.

32. Cheung MWL (2014): Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychol Methods*. 19: 211–229.

33. Viechtbauer W, Cheung MW-L (2010): Outlier and influence diagnostics for meta-analysis. *Res Synth Methods*. 1: 112–125.

34. Rosenthal R (1979): The file drawer problem and tolerance for null results. *Psychol Bull*. 86: 638–641.

35. Duval S, Tweedie R (2000): Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 56: 455–463.

36. Hooijmans CR, Rovers MM, Vries RBM De, Leenaars M, Ritskes-hoitinga M, Langendam MW (2014): SYRCLE ' s risk of bias tool for animal studies. *BMC Med Res Methodol*. 14: 1–9.

37. Assink M, Wibbelink CJM (2016): Fitting three-level meta-analytic models in R: A step-by-step tutorial. *Quant Methods Psychol*. 12: 154–174.

38. Hastie T, Friedman J, Tibshirani R (2009): *The elements of statistical learning: Data Mining, Inference and Prediction*, Second Edi. Springer Series in Statistics. Retrieved from http://link.springer.com/10.1007/978-0-387-21606-5_7.

39. Walker C-D, Bath KG, Joels M, Korosi A, Larauche M, Lucassen PJ, *et al.* (2017): Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress*. 0: 1–28.

40. Loi M, Mossink JCL, Meerhoff GF, Den Blaauwen JL, Lucassen PJ, Joëls M (2017): Effects of early-life stress on cognitive function and hippocampal structure in female rodents. *Neuroscience*. 342: 101–119.

41. Karp NA (2018): Reproducible preclinical research—Is embracing variability the answer? *PLoS Biol*. 16: 1–5.

42. R Core Team (2015): R: A language and environment for statistical computing. *R Found Stat Comput*. . Retrieved from https://www.r-project.org/.

43. Lissa MCJ Van (2018): Package ' metaforest .' *Cran*. .

44. Chang W, Cheng J, Allaire J, Xie Y, McPherson J (2017): shiny: Web Application Framework for R. .

Retrieved from https://cran.r-project.org/package-shiny.

45. Wickham H, François R, Henry L, Müller K, RSudio (2018): Package ' dplyr ' Version 0.7.6. . Retrieved from https://dplyr.tidyverse.org/.

46. Walker AK, Nakamura T, Byrne RJ, Naicker S, Tynan RJ, Hunter M, Hodgson DM (2009): Neonatal lipopolysaccharide and adult stress exposure predisposes rats to anxiety-like behaviour and blunted corticosterone responses: Implications for the double-hit hypothesis. *Psychoneuroendocrinology*. 34: 1515–1525.

47. Pechtel P, Pizzagalli DA (2011): Effects of early life stress on cognitive and affective function: An integrated review of human literature. *Psychopharmacology (Berl)*. 214: 55–70.

48. Suor JH, Sturge-Apple ML, Davies PT, Cicchetti D, Manning LG (2015): Tracing Differential Pathways of Risk: Associations Among Family Adversity, Cortisol, and Cognitive Functioning in Childhood. *Child Dev*. 86: 1142–1158.

49. Hooijmans CR, Ritskes-Hoitinga M (2013): Progress in using systematic reviews of animal studies to improve translational research. *PLoS Med*. 10: 1–4.

50. Rosenthal R (1991): *Meta-analytic procedures for social research.* Sage.

51. Pryce CR, Bettschen D, Nanz-Bahr NI, Feldon J (2003): Comparison of the effects of early handling and early deprivation on conditioned stimulus, context, and spatial learning and memory in adult rats. *Behav Neurosci*. 117: 883–893.

52. Penke Z, Felszeghy K, Fernette B, Sage D, Nyakas C, Burlet A (2001): Postnatal maternal deprivation produces long-lasting modifications of the stress response, feeding and stress-related behaviour in the rat. *Eur J Neurosci*. 14: 747–755.

53. Veenema AH, Bredewold R, Neumann ID (2007): Opposite effects of maternal separation on intermale and maternal aggression in C57BL/6 mice: Link to hypothalamic vasopressin and oxytocin immunoreactivity. *Psychoneuroendocrinology*. 32: 437–450.

54. Antonic A, Sena ES, Lees JS, Wills TE, Skeers P, Batchelor PE, *et al.* (2013): Stem Cell Transplantation in Traumatic Spinal Cord Injury: A Systematic Review and Meta-Analysis of Animal Studies. *PLoS Biol*. 11. doi: 10.1371/journal.pbio.1001738.

55. Egan KJ, Sena ES, Vesterinen HM (2011): Making the most of animal data - Improving the prospect of success in pragmatic trials in the neurosciences. *Trials*. 12: no pagination.

56. Cooper H, Hedges L V, Valentine JC (2009): *The handbook of research synthesis and meta-analysis*

2nd edition, 2nd ed. Russel Sage Foundation.

57. Hensch TK (2005): Critical period plasticity in local cortical circuits. *Nat Rev Neurosci*. 6: 877–888.

58. Levine S, Lewis GW (1959): Critical period for effects of infantile experience on maturation of stress response. *Science (80-)*. 129: 42–43.

59. Ellenbroek BA, Cools AR (2002): Early maternal deprivation and prepulse inhibition: The role of the postdeprivation environment. *Pharmacol Biochem Behav*. 73: 177–184.

60. Witek-Janusek L (1988): Pituitary-adrenal response to bacterial endotoxin in developing rats. *Am J Physiol*. 255: E525–E530.

61. Van Oers HJ j., De Kloet ER, Levine S (1998): Early vs. late maternal deprivation differentially alters the endocrine and hypothalamic responses to stress. *Dev Brain Res*. 111: 245–252.

62. Peña CJ, Kronman HG, Walker DM, Cates HM, Rosemary C, Purushothaman I, *et al.* (2017): Early life stress confers lifelong stress scsceptibility in mice via ventral tegmental area OTX2. *Science (80-)*. 1: 1185–1188.

63. Singh-Taylor A, Molet J, Jiang S, Korosi A, Bolton JL, Noam Y, *et al.* (2018): NRSF-dependent epigenetic mechanisms contribute to programming of stress-sensitive neurons by neonatal experience, promoting resilience. *Mol Psychiatry*. 23: 648–657.

64. Suchecki D (2018): Maternal regulation of the infant's hypothalamic-pituitary-adrenal axis stress response: Seymour 'Gig' Levine's legacy to neuroendocrinology. *J Neuroendocrinol*. 30: 1–17.

65. Rice CJ, Sandman CA, Lenjavi MR, Baram TZ (2008): A Novel Mouse Model for Acute and Long-Lasting Consequences of Early Life Stress. *Endocrinology*. 149: 4892–4900.

66. Teicher MH, Samson JA (2013): Childhood maltreatment and psychopathology: A case for ecophenotypic variants as clinically and neurobiologically distinct subtypes. *Am J Psychiatry*. 170: 1114–1133.

67. Baram TZ, Davis EP, Obenaus A, Sandman CA, Small SL, Solodkin A, Stern H (2012): Fragmentation and unpredictability of early-life experience in mental disorders. *Am J Psychiatry*. 169: 907–915.

68. Whipple EE, Webster-Stratton C (1991): The role of parental stress in physically abusive families. *Child Abuse Negl*. 15: 279–291.

69. Cohen N (2001): *Language impairment and psychopathology in infants, children, and adolescnets*. Sage Publications.

70. Miller AB, Sheridan MA, Hanson JL, McLaughlin KA, Bates JE, Lansford JE, *et al.* (2018): Dimensions of deprivation and threat, psychopathology, and potential mediators: A multi-year longitudinal analysis. *J Abnorm Psychol*. 127: 160–170.

71. Bale TL, Baram TZ, Brown AS, Goldstein JM, Insel TR, McCarthy MM, *et al.* (2010): Early life programming and neurodevelopmental disorders. *Biol Psychiatry*. 68: 314–319.

72. Loi M, Mossink JCL, Meerhoff GF, Den Blaauwen JL, Lucassen PJ, Joëls M (2017): Effects of early-life stress on cognitive function and hippocampal structure in female rodents. *Neuroscience*. 342: 101–119.

73. Riecher-Rössler A (2017): Sex and gender differences in mental disorders. *The Lancet Psychiatry*. 4: 8–9.

74. Plotsky P, Bradley C, Anand K (2000): Behavioral and neuroendocrine consequences of neonatal stress. *Pain Res Clin Manag*. 10: 77–100.

75. Bale TL, Epperson CN (2015): Sex differences and stress across the lifespan. *Nat Neurosci*. 18: 1413–20.

76. Oomen CA, Girardi CEN, Cahyadi R, Verbeek EC, Krugers H, Joëls M, Lucassen PJ (2009): Opposite effects of early maternal deprivation on neurogenesis in male versus female rats. *PLoS One*. 4. doi: 10.1371/journal.pone.0003675.

77. van Hasselt FN, Boudewijns ZSRM, Van Der Knaap NJF, Krugers HJ, Joëls M (2012): Maternal Care Received by Individual Pups Correlates with Adult CA1 Dendritic Morphology and Synaptic Plasticity in a Sex-Dependent Manner. *J Neuroendocrinol*. 24: 331–340.

78. McCarthy MM (2016): Multifaceted origins of sex differences in the brain. *Philos Trans R Soc B Biol Sci*. 371. doi: 10.1098/rstb.2015.0106.