Measuring endogenous corticosterone in laboratory mice - a mapping review, meta-analysis, and open source database

van der Mierden, Stevie ; Leenaars, Cathalijn H C ; Boyle, Erin ; Ripoli, Florenza L ; Gass, Peter ; Durst, Mattea ; Goerlich-Jansson, Vivian C ; Jirkof, Paulin ; Bleich, Andre

DOI: https://doi.org/10.14573/altex.2004221

Review Article

Measuring Endogenous Corticosterone in Laboratory Mice – a Mapping Review, Meta-Analysis, and Open Source Database

Stevie van der Mierden[1], Cathalijn H. C. Leenaars[1], Erin C. Boyle[1], Florenza L. Ripoli[1], Peter Gass[2], Mattea Durst[3], Vivian C. Goerlich-Jansson[4], Paulin Jirkof[3,5], Lydia M. Keubler[1], Steven R. Talbot[1], Anne Habedank[6], Lars Lewejohann[6,7], René H. Tolba[8] and André Bleich[1]

[1]Institute for Laboratory Animal Science, Hannover Medical School, Hannover, Germany; [2]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, University of Heidelberg, Mannheim Faculty, Mannheim, Germany; [3]Division of Surgical Research, University Hospital Zurich, University of Zurich, Zurich, Switzerland; [4]Division of Animals in Science and Society, Utrecht University, Utrecht, The Netherlands; [5]Department of Animal Welfare and 3R, University of Zurich, Zurich, Switzerland; [6]German Federal Institute for Risk Assessment (BfR), German Centre for the Protection of Laboratory Animals (Bf3R), Berlin, Germany; [7]Institute of Animal Welfare, Animal Behaviour and Laboratory Animal Science, Freie Universität Berlin, Berlin, Germany; [8]Institute for Laboratory Animal Science and Experimental Surgery, RWTH Aachen University, Aachen, Germany

Abstract

Evaluating stress in laboratory animals is a key principle in animal welfare. Measuring corticosterone is a common method to assess stress in laboratory mice. There are, however, numerous methods to measure glucocorticoids with differences in sample matrix (e.g., plasma, urine) and quantification techniques (e.g., enzyme immunoassay or radioimmunoassay). Here, the authors present a mapping review and a searchable database, giving a complete overview of all studies measuring endogenous corticosterone in mice up to February 2018. For each study, information was recorded regarding mouse strain and sex; corticosterone sample matrix and quantification technique; and whether the study covered the research theme animal welfare, neuroscience, stress, inflammation, or pain (the themes of specific interest in our consortium). Using all database entries for the year 2012, an exploratory meta-regression was performed to determine the effect of predictors on basal corticosterone concentrations. Seventy-five studies were included using the predictors sex, time-since-lights-on, sample matrix, quantification technique, age of the mice, and type of control. Sex, time-since-lights-on, and type of control significantly affected basal corticosterone concentrations. The resulting database can be used, *inter alia*, for preventing unnecessary duplication of experiments, identifying knowledge gaps, and standardizing or heterogenizing methodologies. These results will help plan more efficient and valid experiments in the future and can answer new questions *in silico* using meta-analyses.

1 Introduction

Glucocorticoids are an important group of steroids that have multiple functions in mammals, including glucose metabolism and anti-inflammatory responses (Ralph and Tilbrook, 2016; Spiga et al., 2011). In animal sciences, the glucocorticoids corticosterone and cortisol have long been measured as indicators of stress (Jones et al., 1998; Palme, 2019; Ralph and Tilbrook, 2016; Newsom and Darrach, 1955), with corticosterone being the main glucocorticoid in Murinae (Spackman and Riley, 1978). Corticosterone secretion is part of the hypothalamus-pituitary-adrenal (HPA) axis, which consists of the hypothalamic paraventricular nucleus (HPN), the pituitary gland, and the adrenal cortex, which releases corticosterone into the bloodstream under the influence of adrenocorticotropic hormone (ACTH) from the pituitary gland. Corticosterone has an inhibitory effect on the HPN, creating a negative feedback loop within the HPA axis (Spiga et al., 2011).

There are multiple factors inherent to the organism that can influence baseline corticosterone concentrations including, but not limited to, sex, age, genetic background, circadian rhythm, and ultradian rhythm (Spiga et al., 2011; Windle et al., 1998; Jones et al., 1998; Spencer and Deak, 2017). A circadian rhythm is an approximately 24-hour cycle in physiological processes (Spen-

Received April 22, 2020; Accepted September 18, 2020;
Epub October 6, 2020; © The Authors, 2020.

ALTEX 38(1), 111-122. doi:10.14573/altex.2004221

Correspondence: André Bleich, PhD
Institute for Laboratory Animal Science, MHH
Carl-Neuberg-Straße 1, 30625 Hannover, Germany
(Bleich.Andre@mh-hannover.de)

cer and Deak, 2017). Corticosterone concentrations follow a circadian rhythm with the peak concentration occurring at the beginning of the active phase (e.g., in the morning for humans and the evening for nocturnal animals like mice and rats) (Spiga et al., 2011). Ultradian rhythms are shorter biological cycles recurring within 24 hours. In humans and rats, the corticosteroid ultradian rhythm has a cycle of approximately one hour (Spiga et al., 2011). To the authors' knowledge, no studies have explicitly measured the ultradian rhythm in mice. The frequency of the ultradian rhythm is consistent over a 24-hour cycle, but the amplitude of corticosterone secretion varies, with lower amplitudes during the inactive phase and higher amplitudes during the active phase (Windle et al., 1998).

Corticosterone concentrations are also affected by external factors like exposure to stressors such as restraint stress or housing conditions, as well as by some types of anesthesia (Spencer and Deak, 2017; Valentine et al., 2012; Jacobsen et al., 2012). Corticosterone is used as a stress indicator in animal sciences as it is relatively easy to measure and generally responds predictably to different types and intensities of stressors (Armario et al., 1986; Anisman et al., 2001). However, the corticosterone response to a particular stressor can vary depending on, for example, sex (Jones et al., 1998), strain (Anisman et al., 2001), age (Foilb et al., 2011), and whether the stressor occurs during the rising or the falling phase of an ultradian cycle (Windle et al., 1998).

Circulating and tissue levels of corticosterone can be measured in animals. Measurements in blood-based matrices like serum or plasma are the most common. A disadvantage of blood sampling is that drawing blood from an animal is stressful and thus might influence the measured corticosterone concentration. Even handling the mouse before drawing blood can activate the HPA axis (Benedetti et al., 2012) unless the time between opening the cage and drawing blood is short enough (Spencer and Deak, 2017). With increasing efforts to reduce distress for laboratory animals, alternative and less invasive methods for determining corticosterone levels have been developed. For example, corticosterone metabolites can be measured in feces and urine (Palme, 2019; Palme et al., 2005). This has several advantages: It is less invasive than blood-based methods, it is less sensitive to contamination by sampling-induced stress, and it is less sensitive to circadian influence (Sheriff et al., 2011). However, this approach has its own challenges, including possible sample contamination, and the potential need to house animals individually, which can influence corticosterone levels (Laber et al., 2008). These various factors affecting corticosterone as well as different experimental setups and quantification techniques can make direct comparisons difficult.

To allow for meaningful generalizations and comparisons between studies, information on the experimental methods is necessary. Systematic mapping reviews are useful tools to synthesize such information. The current work is a mapping review that aimed to identify all studies measuring endogenous corticosterone in mice. The review aimed to answer two main research questions: (i) which sample matrices and methods of detection are used for corticosterone measurement in mice, and (ii) which fields of research (animal welfare, inflammation, neuroscience, pain and/or stress, the fields the R2N consortium focusses on) are the studies measuring corticosterone in mice from? The results give an overview of the current and historic state of research measuring endogenous corticosterone in mice. More importantly, we provide a free, searchable online database of all the relevant papers and the extracted (meta-)data. As an example of how the database can be used, a meta-analysis of a subset of the included studies was performed to determine factors influencing baseline corticosterone concentrations. This database can be used to prevent unnecessary duplication of experiments, identify knowledge gaps, standardize or heterogenize methodologies, and plan future experiments.

2 Methods

2.1 Mapping review protocol

A mapping review protocol was established and published on 2018-02-23 on the Open Science Framework (OSF) website[1] before starting the screening phase. To improve retrievability, the protocol was also published on the Systematic Review Facility[2] on 2019-01-14, and a peer-reviewed version was recently published (Leenaars et al., 2020b).

2.2 Search strategy

For this mapping review, Embase and PubMed databases were searched. The search strategy consisted of two components: corticosterone and mice. Both thesaurus terms (MeSH for PubMed, EMTREE for EMBASE) and title/abstract/keyword searches were used for each component. For the "corticosterone" component, synonyms and truncations with wildcards were identified to retrieve as many relevant papers as possible. For the "mice" component, the search strategy was adapted from the widely-used SYRCLE search strings for PubMed (Hooijmans et al., 2010) and Embase (de Vries et al., 2011). The complete search strategy can be found in the posted protocol and accompanying publication (Leenaars et al., 2020b).

2.3 Study selection

The PubMed and Embase searches were performed on 2018-02-07. Duplicate entries were removed manually using Endnote X8 (Clarivate). References were then uploaded to Early Review Organizing Software (EROS)[3] for the inclusion and exclusion of references based on the *a priori* criteria defined in the protocol. The inclusion criteria were: (i) reference must comprise a primary study, (ii) study was performed in house mice (*Mus musculus*), and (iii) endogenous corticosterone was measured. Studies in which, e.g., corticosterone was administered without measuring a baseline were thus excluded. Studies only reporting measurements of one or more corticosterone metabolites were excluded;

[1] https://osf.io/8yt3b

[2] http://syrf.org.uk/

[3] https://www.eros-systematic-review.org

this included studies having measured corticosterone in feces and urine, as corticosterone is heavily metabolized and very little to no corticosterone remains in these matrices (Touma et al., 2003).

Because of the broad criteria and large number of papers, the title/abstract screening phase and the full-text screening phase were combined into one phase. Reviewers first tried to include or exclude a reference based on the title, abstract, and keywords. If they could not decide based on this alone, they went to the full text for a definitive inclusion or exclusion. Seven reviewers assisted with the screening (FLR, ECB, SvdM, MD, PJ, LMK, CHCL). All references were screened independently by at least two reviewers. Discrepancies were discussed by at least two reviewers until consensus was reached.

2.4 Data extraction

References were distributed among different reviewers. Data was extracted using a standardized Excel spreadsheet[4]. To minimize variation between reviewers, pre-defined options were used when possible (e.g., for sex, the options were "M" for male, "F" for female, "B" for both, or "U" for unknown). Of all the references extracted by a reviewer, at least 5% were randomly checked by a second reviewer for errors, and all were checked for consistency (i.e., that all adhered to the same format). In total, eleven reviewers performed the data extraction (AH, BS, ECB, LMK, LL, MD, PG, PJ, RK, SvdM, and VCGJ). The authors extracted the following data: i) study identification information (authors, journal, year of publication, etc.); ii) mouse strain and sex; iii) sample matrix wherein corticosterone was measured (e.g., plasma); iv) whether the mice modelled a human disease; v) corticosterone quantification method; and vi) whether the article dealt with animal welfare, inflammation, neuroscience, pain, and/or stress. Given the scope of this review, if information was referenced in the included paper, we did not retrieve the referenced paper, but noted "referenced". Conference abstracts, posters, etc. were excluded from the results unless explicitly stated otherwise.

A crossing between different mouse strains was indicated by an ampersand (e.g., C57BL/6 & C3H). If the study mentioned the use of a vendor strain, it is also presented in the "mixed vendor strain" column in the file[4], in this case as B6C3F1/J (The Jackson Laboratory). This separation makes it possible to specifically identify studies using either vendor strains (e.g., B6C3F1/J) or strain crossings with the strains of interest (e.g., all hybrid strains with C57BL/6 or C3H strains).

A study was identified as modelling a human disease when the authors of that paper mentioned that the mice were used for modelling (part of) a human disease.

2.5 Meta-regression

Data extraction

The protocol for the meta-regression (MR) was not preregistered. All papers from 2012 were selected as a sample. This selection was made before any additional data were extracted.

The following data were additionally extracted for the meta-analysis: i) mean age of mice in weeks at the time of corticosterone measurement; ii) time in hours between lights-on (time 0) and corticosterone measurement ("time-since-lights-on"); iii) mean housing temperature of the mice in Celsius; iv) whether mice were anaesthetized when the corticosterone measurement was taken; v) whether the ultradian rhythm was considered when measuring the corticosterone concentration, and vi) whether the control group was intervention-free (designated "naïve control") or whether it underwent a treatment like a sham operation or vehicle injection (designated "sham control"). When any of this information was not reported in the paper, efforts were made to contact the authors of the original paper to request it.

The corticosterone concentration, mean age of the mice, and the mean temperature at which the mice were kept are continuous variables. Anesthesia, ultradian rhythm, and controls are dichotomous variables. Time-since-lights-on can range from 0 to 24 hours, yet because of the cyclical nature of the circadian rhythm, the expected corticosterone concentration at, for example, 23 hours time-since-lights-on is very close to the expected concentration at 0 hours time-since-lights-on. Therefore, time-since-lights-on was converted to a categorical variable consisting of four groups: i) "lights on" encompassing all concentrations measured between 22-2 hours; ii) "day period" encompassing 2-10 hours; iii) "lights off" encompassing 10-14 hours; and iv) "night period" encompassing 14-22 hours (Fig. S1[5]).

Corticosterone concentrations were converted to ng/mL where necessary and possible. Standard errors of the mean (SEM) and confidence intervals were converted to the standard deviation (SD) where necessary and possible. If the concentration was only presented graphically, the graphics editing software GIMP 2.0 was used to determine the concentration from the graph based on the number of pixels. Additionally, the authors of the original paper were contacted to request the exact concentrations and deviations.

Study inclusion

Not all studies reported all data of interest. To prevent missing data from decreasing the power of the MR, a *post-hoc* selection of predictors was made after visual inspection of the data. To be included in the MR, studies also needed to report the corticosterone concentration so that it could be converted to ng/mL (see above). The predictors analyzed in the first MR were sex (male/female/both), time-since-lights-on (lights on/day period/lights off/night period), sample matrix (plasma/serum), quantification technique (HPLC/"EIA/ELISA"/RIA), and control type (naïve control/sham control). A second MR was performed to analyze the effect of mouse strain. This predictor could not be included in the primary MR because many strains were used in only a few studies. Only strains for which there were at least five observations were included in this second MR. As a result of the *post-hoc* predictor selection and sample choice, this MR should be interpreted as an exploratory test only.

[4] doi:10.14573/altex.2004221s2

[5] doi:10.14573/altex.2004221s1

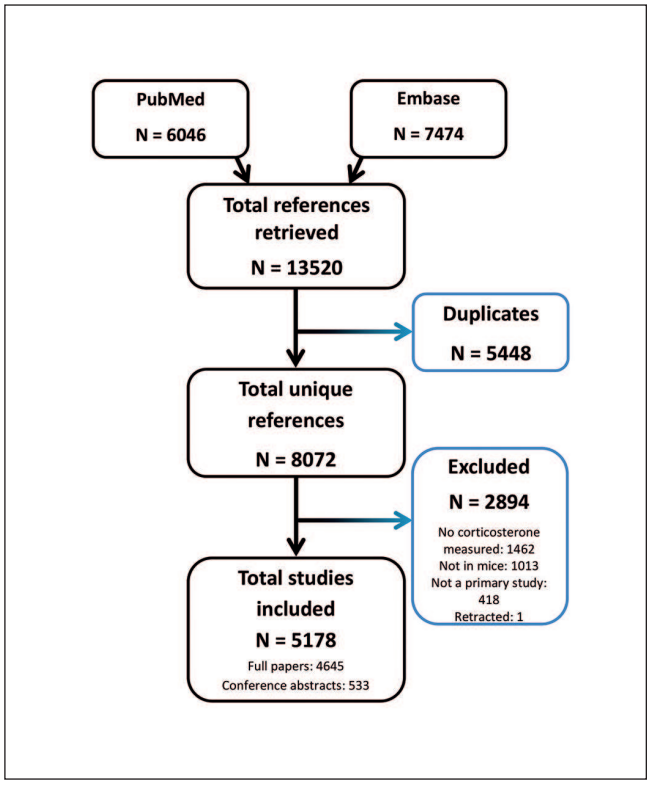

Fig. 1: Flow scheme of the references included in the mapping review

Of the original 13,520 retrieved references, 5178 were included for data extraction. 533 articles were conference abstracts. Conference abstracts were excluded from all other graphs and tests.

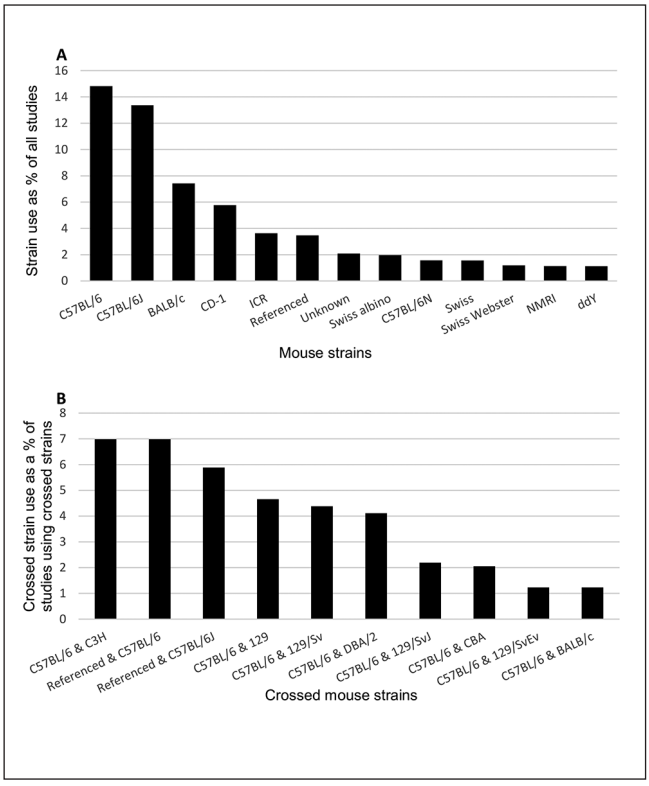

Fig. 2: The most common mouse strains (A) and crossed mouse strains (B) used in studies measuring endogenous corticosterone

Analysis

MR was performed in R (R Core Team, 2019) using the metafor package (Viechtbauer, 2010). The rma.uni function was used with the Knapp-Hartung modification. An omnibus test was performed for the sex, time-since-lights-on, and quantification technique moderators with the ANOVA.rma function. Because of the large spread and skew of the data, a transformation of the concentrations and corresponding standard deviations as proposed by Higgins et al. (2008) was performed. A sensitivity analysis was performed – excluding outliers – to confirm the result of the primary MR.

3 Results

3.1 Retrieval of all studies measuring endogenous corticosterone in mice

A systematic search for all studies measuring endogenous corticosterone in mice was undertaken. The flow diagram presents the number of references in each phase of the review (Fig. 1). The initial search retrieved 13,520 references. Of these, 5,448 were duplicate records, leaving 8,072 unique references to be screened for in- or exclusion. 2,894 references were excluded based on the reasons presented in Figure 1. 5,178 references were included for data extraction. 533 (10.3%) of the 5,178 references were identified as conference abstracts, while 4,645 references were full-length papers. A list of all included papers is available in a searchable database[4].

For 324 (7%) of the 4,645 full-length papers, the full text could not be retrieved. In most cases the abstract was available. The title and abstract were used to extract as much information as possible. To prevent creating a biased representation of reporting frequencies, the papers for which the full text could not be retrieved were excluded from calculations of reporting frequencies, unless specifically stated otherwise.

In general, the annual number of papers measuring endogenous corticosterone in mice has greatly increased – from one publication in 1955 to 351 papers in 2017 (Fig. S2[4]), the last complete year included in the mapping review.

3.2 Corticosterone is measured most often in male C57BL/6 mice

From the mapping review database, one can acquire information on the use of different mouse strains. From the 4,645 full text papers, 432 single mouse (sub)strains were identified. However, it is likely that the actual number of strains used differs, as the re-

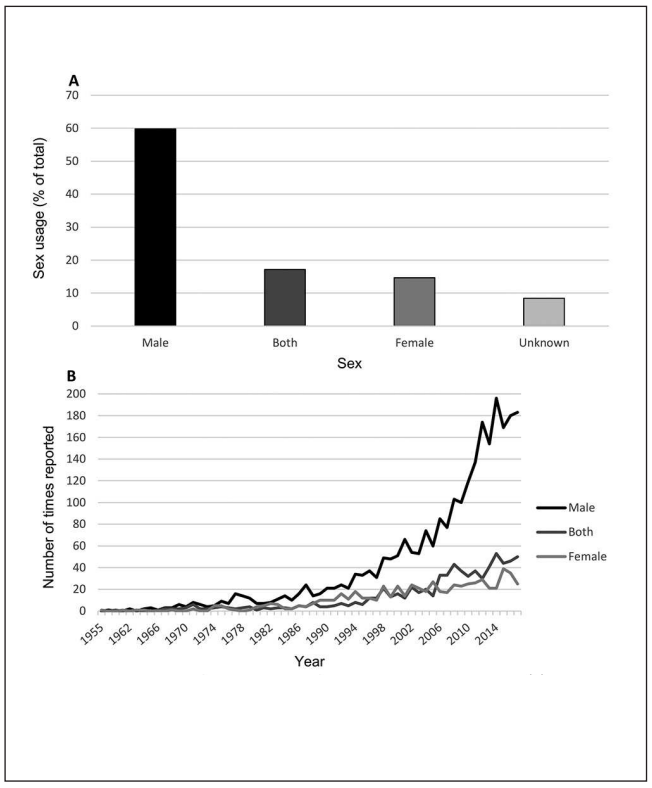

Fig. 3: Reported mouse sexes used in corticosterone measurements (A) overall and (B) over time

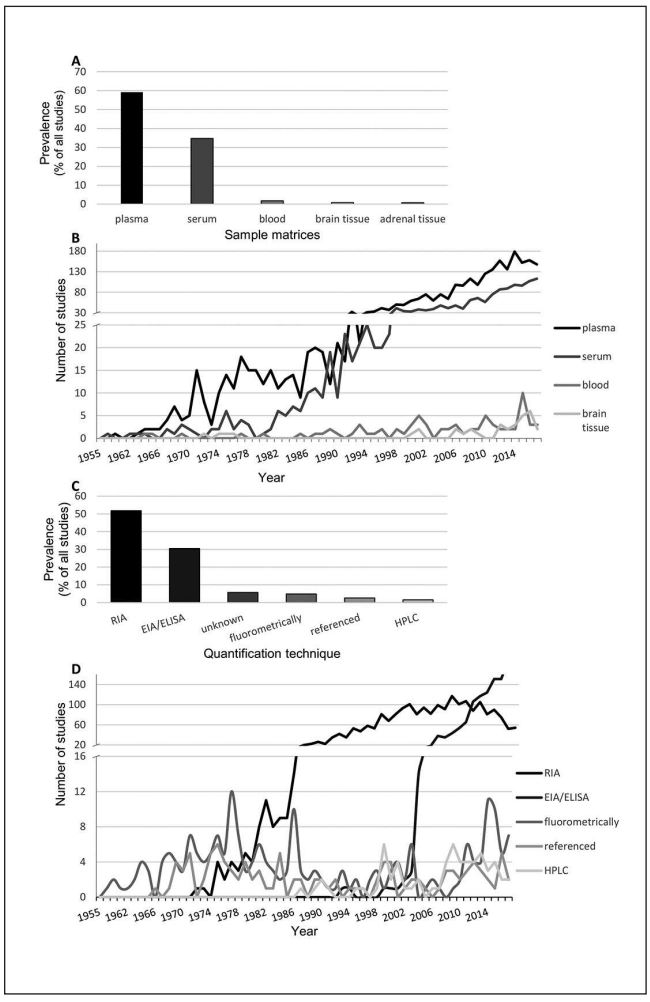

Fig. 4: The total use of different sample matrices for measuring corticosterone in mice (A), the sample matrices used over time (B), the total use of different quantification techniques (C), and the quantification techniques used over time (D)

"Referenced" indicates studies not reporting the quantification technique, but referencing another study for the information.

porting of strains was inconsistent. All reported mouse strains are presented in the database[4].

Figure 2A shows the most frequently reported mouse strains used for corticosterone measurements. Each of these strains was used in at least 1% of all included papers. "C57BL/6" is the mouse strain most commonly reported, followed by C57BL/6J, with the "J" specifying the specific C57BL/6 substrain from the Jackson Laboratory. Together, C57BL/6(J) mice were used in over a quarter of all studies. In contrast, many other strains were used in only a few studies, with 418 strains used in less than 1% of the included papers. Interestingly, the sixth-largest group is the "referenced" category (3.4% of all included papers referred to another paper for information on the tested strain), in which the papers do not directly state the mouse strain, but rather refer to other papers. The seventh-largest group is the "unknown" (2.1%) category, in which the papers mention neither the strain nor refer to other papers.

The ten most common crossed mouse strains used for corticosterone measurements used C57BL/6 as at least one of the parental strains (Fig. 2B). A cross between C57BL/6 & C3H was used in 7.0% of all papers using crossed strains (51 papers). Some strain crossings are available from vendors (e.g., B6C3F1/J hybrid mice (The Jackson Laboratory)), however, quite often the source of crossed mice was not stated.

From the database, one can also conclude that corticosterone was most often measured in male mice, with almost 60% of the papers using male mice only (Fig. 3). There were more studies that used both male and female mice (17.1%) than there were studies that only used female mice (14.7%). 8.4% of the papers did not specify which sex was used.

3.3 Endogenous corticosterone is primarily measured in plasma and serum

One of the primary objectives of this mapping review was to determine which sample matrices were used for corticosterone measurement in mice. From all the studies measuring endogenous corticosterone in mice, 45 different sample matrices were identified[4]. Figure 4A shows the ten most used sample matrices.

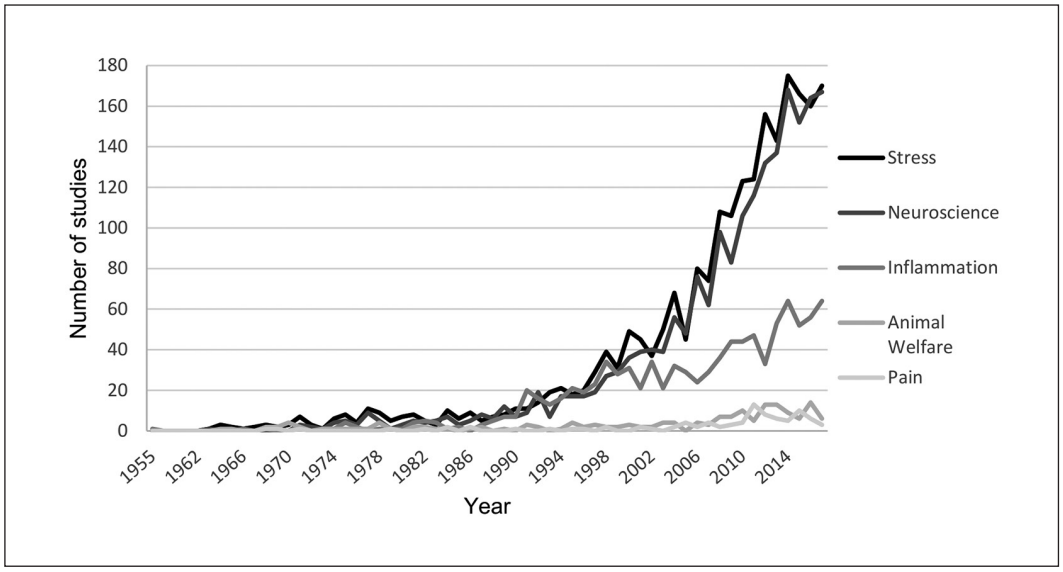

Fig. 5: Number of studies related to the five specific research themes over time

Together, plasma and serum make up over 93% of all reported sample matrices. The reporting of sample matrix was high, with only 0.63% of the included papers not reporting in which matrix corticosterone was measured.

The use of different sample matrices is presented over time in Figure 4B. Plasma and serum have remained the most common sample matrices since corticosterone measurements began.

3.4 Corticosterone is most frequently quantified using immunoassays

There are various techniques to quantify specific steroid hormones in biological samples. Corticosterone was measured in mice using 32 different quantification techniques[4]. Figure 4C shows the quantification techniques reported in at least 1% of all studies. Radioimmunoassay (RIA) was the most common method, being used in over 50% of the papers, followed by enzyme immunoassays (EIA/ELISA) (30.9%). The third largest group was the "unknown" group: 5.7% of all included papers did not report the corticosterone quantification technique used.

The relative use of quantification techniques changed over time (Fig. 4D). Initially, fluorometric techniques were the most common, but they were succeeded by RIA. More recently, enzyme immunoassay-based techniques have gained popularity.

3.5 Corticosterone is most often measured in the context of neuroscience and stress

This mapping review was also interested in determining in which fields of research corticosterone was being measured. Note that a study can comprise different research themes and is then included multiple times in this analysis. Relatively few papers studied corticosterone in the context of pain (91 papers, 2.1%) or animal welfare (164 papers, 3.8%). Neuroscience-related studies (2010 papers, 46.5%) and stress-related studies (2277 papers, 52.7%) were far more common. There were 971 inflammation-related studies (22.3%).

Figure 5 shows how the number of papers relating to the different research themes changed over time. The graph shows the number of papers relating to stress or neuroscience increasing, while such a trend seems absent or strongly reduced for animal welfare-, pain-, or inflammation-related papers. Especially for animal welfare, there may be more studies that measure only corticosterone metabolites and are therefore not included in this review. The number of studies measuring corticosterone metabolites in feces seems to be increasing (Palme, 2019).

3.6 Study subset selection for MR

As an example of how the database could be used, a meta-analysis of a subset of the included studies was performed to determine factors influencing baseline corticosterone concentrations. Using all database entries for the year 2012, 265 studies were initially identified for inclusion in the MR. Several of these studies were excluded from the analysis as they did not present corticosterone concentrations of an intervention-free group or because they presented the concentration in a unit that could not be converted to ng/mL (e.g., ng/g for corticosterone in adrenal tissue). Altogether, 196 papers reported a usable baseline corticosterone concentration, as well as either the standard deviation or standard error of the mean, and the number of mice per group (N). Studies that did not report the selected predictors were also removed from the MR. This left 75 studies with 106 measurements in total (Tab. S1[5]).

The spread of the reported baseline corticosterone concentrations was large (Fig. 6). The lowest reported concentration was $7x10^{-3}$ ng/mL and the highest was $2.04x10^{6}$ ng/mL – a difference of approximately eight orders of magnitude. A histogram of the reported corticosterone concentrations shows that the data was skewed, so a Higgins transformation was performed to normalize the distribution of the concentration data (Fig. S3[5]). Only a few low-concentration outliers remained, but based on the papers that these data points were derived from, there were no reasons to exclude these concentrations from the MR.

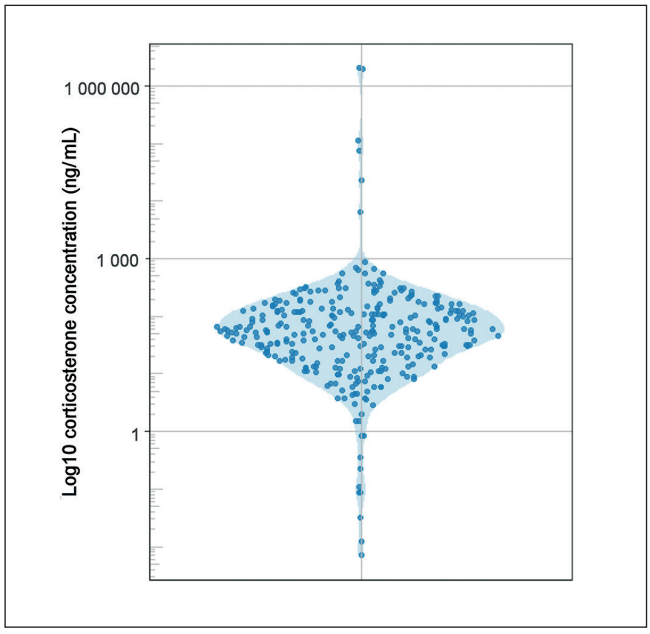

Fig. 6: Spread of all reported corticosterone concentrations on a y log-axis
The highest reported concentration was 2.04x10^{6} and the lowest was 7x10^{-3} ng/mL.

3.7 MR revealed an effect of sex, time-since-lights-on, and type of control on basal corticosterone concentration

An MR was performed on the 75 included studies from 2012 to determine the effect of predictors on basal corticosterone concentrations. The results of the MR are presented in Table 1. The omnibus tests showed a significant effect of sex and the time-since-lights-on moderators. Female mice had significantly higher corticosterone concentrations than male mice ($p < 0.001$, effect size estimate: 1.65, CI: 0.76-2.54, Fig. 7A). For the time-since-lights-on, there was a clear difference between the lights-off and lights-on period ($p < 0.01$, effect size estimate: 1.59, CI: 0.49-2.68), with the corticosterone concentration being higher during lights-off (Fig. 7B). Furthermore, the MR showed that the mice which did not undergo sham or vehicle treatments (naïve control) had a lower reported baseline corticosterone concentration than the mice which did ($p < 0.01$, effect size estimate: -1.33, CI: -2.10 – -0.54), Fig. 7C). The overall heterogeneity was high with an I^2 of 99.43%. None of the other tested moderators (sample matrix, quantification technique, or age) showed a significant effect on the reported corticosterone concentrations. The age of the mice ranged from 1 to 39 weeks, with the median at 12 weeks.

The results of the sensitivity analysis, in which the outliers of corticosterone concentrations were removed, were consistent with the main MR (data not shown).

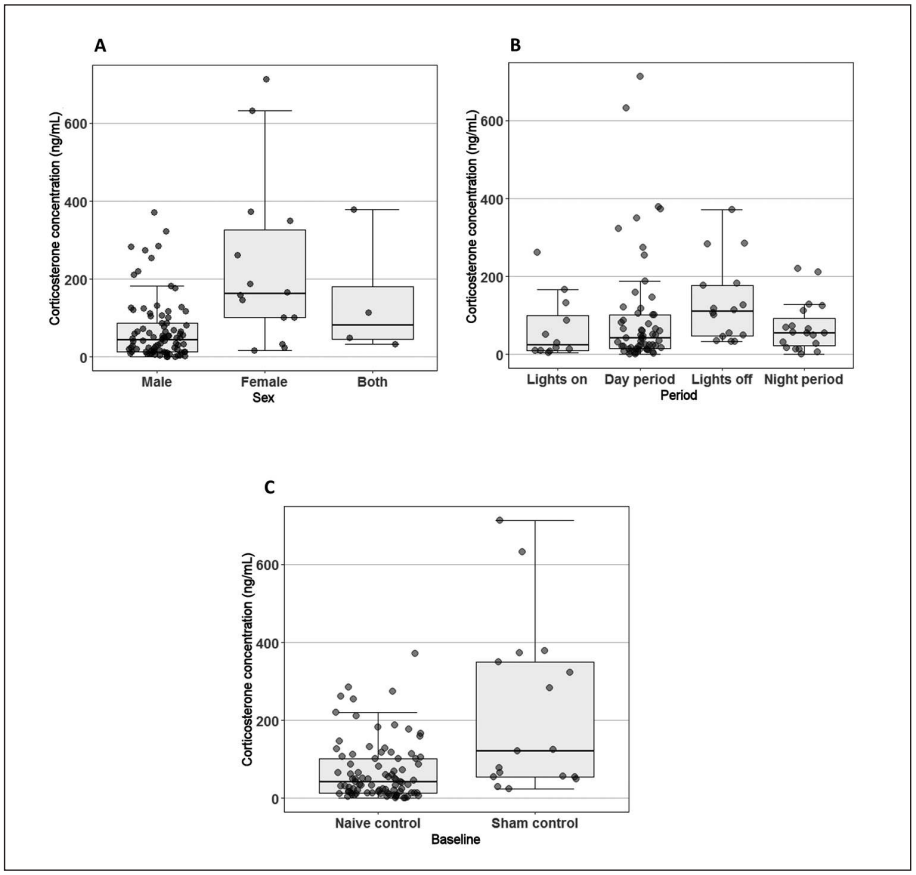

Fig. 7: The corticosterone concentrations of references included in the meta-regression for the predictors that significantly influence the basal corticosterone concentrations: sex of the mice (A), time-since-lights-on (B), and type of control (C)
Each dot represents one data point.

Tab. 1: Meta-regression of the effect of moderators on reported corticosterone concentrations in mice – Results of the omnibus test (A) and meta-regression (B)

A

Moderator	F value (df)	p-value
Sex	7.272 (2.95)	0.001
Time-since-lights-on	5.078 (3.95)	0.003
Quantification technique	0.076 (3.95)	0.973

B

Moderator	Range or values	Estimate from MR (SE)	p-value from MR	95% CI – lower bound	95% CI – upper bound
Sex (C)	Male, female, both				
	Female versus male	1.64 (0.43)	< 0.001	0.78	2.50
	Both versus male	0.38 (0.72)	0.602	-1.06	1.81
Time since lights on (C)	Lights-on, day period, lights-off, night period				
	Day period versus lights-on	0.06 (0.44)	0.900	-0.82	0.93
	Lights-off versus lights-on	1.60 (0.55)	0.004	0.51	2.68
	Night period versus lights-on	0.27 (0.53)	0.613	-0.78	1.31
Sample matrix (C)	Plasma versus serum	0.00 (0.32)	0.992	-0.64	0.64
Quantification technique (C)	HPLC, EIA/ELISA, RIA				
	EIA/ELISA versus HPLC	0.06 (1.02)	0.956	-1.96	2.08
	RIA versus HPLC	0.15 (1.02)	0.882	-1.88	2.19
Age (N)	0-66.5 weeks	-0.01 (0.02)	0.703	-0.05	0.03
Intervention free (C)	Naïve control versus sham control	-1.34 (0.39)	< 0.001	-2.11	-0.57

3.8 MR revealed no effect of mouse strain on basal corticosterone concentration

A second MR was performed to analyze the effect of mouse strain on basal corticosterone concentration. Only five strains were compared in this second MR, since all other mouse strains were described in less than five publications from 2012. Baseline corticosterone concentrations were compared among the strains C57BL/6 (n = 229), BALB/c (n = 15), CD-1 (n = 44), CR (n = 16), and NMRI (n = 18). MR of these strains showed no significant difference in the basal corticosterone concentrations among these strains (Fig. 8).

4 Discussion

Mapping reviews are a useful tool to present an overview of the current state of a research topic (Grant and Booth, 2009). The results of a mapping review can prevent unnecessary duplication of animal experiments and promote the adequate design, reporting, and analysis of experiments. To allow for meaningful generalizations and comparisons between studies, information on the experimental methods is necessary. This review highlights several study characteristics for which reporting could be improved. Improving reporting can help improve the internal and external validity of animal experiments and reduce the number of animals needed, one of the core principles of the three Rs (Reduce, Refine, Replace) (Russell and Burch, 1959).

In this study, we performed a mapping review and created an accessible online database[6] for all studies measuring endogenous corticosterone in mice until February 2018. Large systematic reviews take a long time to complete (e.g., Cochrane reviews take on average 67.3 weeks (Borah et al., 2017), and systematic (mapping) reviews of animal studies are no exception; refer to, e.g., two recently published reviews with search dates in December 2015

[6] doi:10.17605/OSF.IO/XNH24

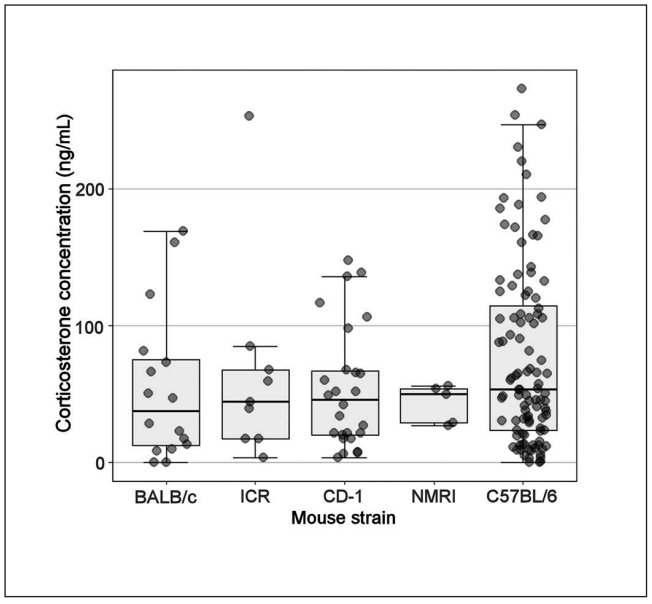

Fig. 8: Corticosterone concentrations for all strains included in the second meta-regression
Each dot represents one data point.

(Leenaars et al., 2020a) and August 2016 (Archer et al., 2018)). As review updates also take time, a lag-time between search and publication is unavoidable. To retrieve more recent studies, the search strategy for this review can be combined with specific terms of interest, and the publication date limited from February 2018 onwards.

We chose to focus on mice as they are the most used animal model in science[7,8]. Corticosterone is an important hormone of interest in many different fields of study including stress, neuroscience, immunology, animal welfare, and endocrinology. This review found few studies using corticosterone measurements in the context of animal welfare. While the analysis of well-being states in animals is increasing, especially for animal-based research (Keubler et al., 2020; Bleich and Tolba, 2017), there is still less research measuring corticosterone in mice in animal welfare than in neuroscience. However, we excluded studies analyzing fecal corticosterone metabolites, a method which potentially can replace measuring corticosterone in serum or plasma (Palme, 2019).

Currently, there is a reproducibility crisis in animal science, and failures in both standardization and heterogenization may be partially responsible. Standardization can improve reproducibility by decreasing between-experiment variation (Beynen et al., 2001), and standardization of biologically irrelevant variables could exclude the effect of this variable on the relevant outcome. For example, our MR showed that the reported control type (i.e., "naïve" control versus sham intervention) influenced the basal corticosterone concentration. This is in line with the literature (Drude et al., 2011). It is important to consider the influence of the control type on the experiment, especially since there is a maximum to the corticosterone secretion (Spencer and Deak, 2017). It is also important to take this potential influence into account when comparing the results of different studies.

On the other hand, overly standardizing potentially relevant biological variables, like sex or strain, can lead to idiosyncratic results (Bailoo et al., 2014). The mapping review showed that the majority of studies were performed in male C57BL/6 mice. Previous studies have shown that the (sub)strain and even the vendor can influence experimental results (Åhlgren and Voikar, 2019; Mulligan et al., 2008; Ashworth et al., 2015; Tuttle et al., 2018;[9]). Furthermore, the MR showed that sex influences baseline corticosterone concentrations – an observation also substantiated in the literature (Sittig et al., 2016; Grad and Khalid, 1968; Caruso et al., 2018). Thus, the choice of sex and strain could influence corticosterone concentration results. For future experiments, authors should therefore consider using both sexes and different or multiple strains when possible and appropriate.

Incomplete reporting of methods also likely contributes to the reproducibility crisis (Kilkenny et al., 2009). This mapping review revealed that reporting is often incomplete, especially regarding the mouse strain lab codes. There was also incomplete reporting of sample matrices: sometimes the terms plasma and serum were used interchangeably, or the term "blood" was used. Directly measuring glucocorticoid concentrations in whole blood with immunoassays is not advised due to possible interference with blood cells (Spencer and Deak, 2017). Accordingly, we presume "blood" was used as shorthand for either plasma or serum. A consideration is that for most methods, glucocorticoids measured in serum or plasma represent the "total glucocorticoids", i.e., free glucocorticoid levels plus globulin-bound glucocorticoids. However, it is assumed that only the free glucocorticoids are active. Yet, the fraction of free glucocorticoids is not static and can vary, *inter alia*, depending on stress (Spencer and Deak, 2017; Sheriff et al., 2011). This means that, for most methods, the measured concentrations do not adequately reflect the biologically active concentrations. During screening, multiple studies were found that reported corticosterone concentrations measured in feces or urine. In these cases, we suspect corticosterone metabolites were actually measured, as these matrices contain little to no corticosterone (Palme, 2019; Palme et al., 2005). Furthermore, quantification techniques were not always specified. Authors should correctly report all relevant methodological information.

The MR and the literature show that the time-since-light-on influences the baseline corticosterone concentrations (Spiga et al., 2011). It is important to consider that mice are nocturnal, and

[7] Bundesministerium für Ernährung und Landwirtschaft (2018). Verwendung von Versuchstieren im Jahr 2018. https://www.bmel.de/DE/themen/tiere/tierschutz/versuchstierzahlen2018.html

[8] UK Home Office (2018). Annual Statistics of Scientific Procedures on Living Animals, Great Britain 2018. URL: https://www.gov.uk/government/statistics/statistics-of-scientific-procedures-on-living-animals-great-britain-2018

[9] https://www.jax.org/news-and-insights/jax-blog/2016/june/there-is-no-such-thing-as-a-b6-mouse

studies performing an intervention during the light period might not represent normal physiological conditions, which is important when translating findings to humans. Since the time-since-lights-on variable influences the baseline corticosterone, it can be argued that the timeframe should be standardized to increase reproducibility. However, by restricting the timeframe, idiosyncratic results may be obtained (e.g., the intervention only has an effect during the onset of the active phase), which might hinder translation of the results. In any case, careful consideration of the timeframe and the consequences of this choice for the reproducibility and translational value of the results should be made.

The MR also found that the range of reported concentrations is very wide. This is not uncommon in pre-clinical studies, as a previous systematic review into cerebral adenosine concentrations found 4 orders of magnitude difference between the lowest and highest reported concentration (van der Mierden et al., 2018). A possible explanation can be reporting in the wrong unit. In this meta-regression the concentrations were reported in several units, including mg/dL and pg/mL, which differ by several orders of magnitude.

Inadequate reporting prohibited the analysis of all desired predictors – especially the possible effects of strain and anesthesia (Jones et al., 1998; Van Loo et al., 2004; Jacobsen et al., 2012; Valentine et al., 2012). The other predictors initially considered for inclusion could also explain (part of) the heterogeneity (I^2) found in the MR, but other, not initially considered, predictors might also be relevant, e.g., single versus group housing (Laber et al., 2008). More complete reporting might therefore not only help alleviate the reproducibility problem but also enable more and better meta-research.

The influence of the ultradian rhythm has been discussed extensively in the literature (Spiga et al., 2011; Spencer and Deak, 2017) but could not be included in the MR because very few primary studies on the ultradian rhythm have been published. Based on the literature, the ultradian rhythm can influence experimental results, (Spiga et al., 2011) but the short cycle may spread these differences over different groups, averaging out its effect if the groups and intervention are properly distributed. It is unlikely that there will be enough studies to do a proper meta-analysis into the effect of the ultradian rhythm in the near future, but it is an important subject for more research.

Historical and current standards can help inform which experimental variables can be standardized or homogenized. By searching the corticosterone database for variables of interest, researchers can adjust their experiment to increase standardization or homogenization. In addition, the database can be used to identify all relevant papers for meta-analyses and filter for other variables of interest. Our MR from all papers published in 2012 shows that sex, control type, and time-since-light-on influenced corticosterone baseline concentrations but found no significant effect of sample matrix, quantification technique or age of the mice. Although there are studies that find that basal corticosterone increases with mouse age (Elias and Redgate, 1975), these results were not found in the meta-regression. The absence of a significant result for a potential effect of age in the MR could be due to a lack of studies using aged mice or due to the large spread of the reported concentrations. Note that non-significant predictors in a meta-regression are not necessarily irrelevant (Borenstein et al., 2015). No interactions were tested in the MR because first, we had no *a priori* expectations of interactions, and second, there were not enough observations to include all possible interactions. As a result of the *post-hoc* predictor choices and the sampling of 2012, this MR is exploratory, and the results should be interpreted with caution and in consideration of all known information.

RIA was initially the most common quantification technique but this has been surpassed by EIA. The MR did not show differences in basal corticosterone levels between different quantification techniques, however, the literature indicates generally high levels of variation in concentrations measured via RIA or ELISA (Bekhbat et al., 2018; Lewis and Elder, 1985; Rød et al., 2017; Fanson et al., 2017). This large variation may account for why no significant differences were found. HPLC and related techniques generally have higher precision and accuracy (Turpeinen and Hämäläinen, 2013; Oka et al., 1987) but are used less frequently. The choice of quantification technique could be based on the research question: If only relative differences are of interest, immunoassays can be used. If absolute values are of interest, HPLC or other chromatography-based techniques like LC-MS should be used, as these methods have a higher specificity (Murtagh et al., 2013). Of note, while relative differences are often sufficient for the scientific question at hand, the reuse of data in meta-research is becoming more important with the increasing focus on the 3Rs. Considering this, absolute (not relative) glucocorticoid concentrations are informative to more research questions.

This mapping review provides an overview of the current and historical state of measuring corticosterone in mice and provides a searchable database of all included studies. The meta-regression indicates which predictors may influence corticosterone concentrations. Based on the results of this study, it is important that future experiments report study details as completely and as clearly as possible, especially regarding the mouse strain. Following the PREPARE (Smith et al., 2018) and ARRIVE (Kilkenny et al., 2010) guidelines can help to improve the reporting of animal experiments. In addition, an appropriate experimental design should be selected, considering the sex of the mice, the time-since-lights-on, and the control condition. This does not mean that one should only use female mice or only perform experiments during lights-off, but these factors need to be balanced over the different interventions when setting up the experiment to allow for unconfounded, meaningful results. Together, this review will help researchers plan experiments related to corticosterone in mice with high external validity.

References

Åhlgren, J. and Voikar, V. (2019). Experiments done in Black-6 mice: What does it mean? *Lab Anim 48*, 171-180. doi:10.1038/s41684-019-0288-8

Anisman, H., Hayley, S., Kelly, O. et al. (2001). Psychogenic, neurogenic, and systemic stressor effects on plasma corticosterone and behavior: Mouse strain-dependent outcomes. *Behav Neurosci 115*, 443-454. doi:10.1037/0735-7044.115.2.443

Archer, D. P., McCann, S. K., Walker, A. M. et al. (2018). Neuroprotection by anaesthetics in rodent models of traumatic brain injury: A systematic review and network meta-analysis. *Br J Anaesth 121*, 1272-1281. doi:10.1016/j.bja.2018.07.024

Armario, A., Lopez-Calderon, A., Jolin, T. et al. (1986). Sensitivity of anterior pituitary hormones to graded levels of psychological stress. *Life Sci 39*, 471-475. doi:10.1016/0024-3205(86)90527-8

Ashworth, A., Bardgett, M., Fowler, J. et al. (2015). Comparison of neurological function in males and females from two substrains of C57BL/6 mice. *Toxics 3*, 1-17. doi:10.3390/toxics3010001

Bailoo, J. D., Reichlin, T. S. and Würbel, H. (2014). Refinement of experimental design and conduct in laboratory animal research. *ILAR J 55*, 383-391. doi:10.1093/ilar/ilu037

Bekhbat, M., Glasper, E. R., Rowson, S. A. et al. (2018). Measuring corticosterone concentrations over a physiological dynamic range in female rats. *Physiol Behav 194*, 73-76. doi:10.1016/j.physbeh.2018.04.033

Benedetti, M., Merino, R., Kusuda, R. et al. (2012). Plasma corticosterone levels in mouse models of pain. *Eur J Pain 16*, 803-815. doi:10.1002/j.1532-2149.2011.00066.x

Beynen, A. C., Gärtner, K. and van Zutphen, L. F. M. (2001). Standardization of animal experimentation. In L. F. M. van Zutphen, V. Baumans and A. C. Beynen (eds.), *Principles of Laboratory Animal Science: A Contribution to the Humane Use and Care of Animals and to the Quality of Experimental Results*. 2nd edtition. Amsterdam, The Netherlands: Elsevier.

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:10.1177/0023677217733010

Borah, R., Brown, A. W., Capers, P. L. et al. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open 7*, e012545. doi:10.1136/bmjopen-2016-012545

Borenstein, M., Hedges, L. V., Higgins, J. P. et al. (2015). Regression in meta-analysis. *Comprehensive Meta Analysis Manual*. Englewood, USA: Biostat.

Caruso, M., Reiss, D., Caulfield, J. et al. (2018). Adolescent chronic variable social stress influences exploratory behavior and nicotine responses in male, but not female, BALB/cJ mice. *Brain Res Bull 138*, 37-49. doi:10.1016/j.brainresbull.2017.08.001

de Vries, R. B., Hooijmans, C. R., Tillema, A. et al. (2011). A search filter for increasing the retrieval of animal studies in Embase. *Lab Anim 45*, 268-270. doi:10.1258/la.2011.011056

Drude, S., Geissler, A., Olfe, J. et al. (2011). Side effects of control treatment can conceal experimental data when studying stress responses to injection and psychological stress in mice. *Lab Anim 40*, 119-128. doi:10.1038/laban0411-119

Elias, P. K. and Redgate, E. (1975). Effects of immobilization stress on open field behavior and plasma corticosterone levels of aging C57BL/6J mice. *Exp Aging Res 1*, 127-135. doi:10.1080/03610737508257954

Fanson, K. V., Németh, Z., Ramenofsky, M. et al. (2017). Inter-laboratory variation in corticosterone measurement: Implications for comparative ecological and evolutionary studies. *Methods Ecol Evol 8*, 1745-1754. doi:10.1111/2041-210X.12821

Foilb, A. R., Lui, P. and Romeo, R. D. (2011). The transformation of hormonal stress responses throughout puberty and adolescence. *J Endocrinol 210*, 391-398. doi:10.1530/JOE-11-0206

Grad, B. and Khalid, R. (1968). Circulating corticosterone levels of young and old, male and female C57B1/6J mice. *J Gerontol 23*, 522-528. doi:10.1093/geronj/23.4.522

Grant, M. J. and Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Info Libr J 26*, 91-108. doi:10.1111/j.1471-1842.2009.00848.x

Higgins, J. P., White, I. R. and Anzures-Cabrera, J. (2008). Meta-analysis of skewed data: Combining results reported on log-transformed or raw scales. *Stat Med 27*, 6072-6092. doi:10.1002/sim.3427

Hooijmans, C. R., Tillema, A., Leenaars, M. et al. (2010). Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. *Lab Anim 44*, 170-175. doi:10.1258/la.2010.009117

Jacobsen, K. R., Kalliokoski, O., Teilmann, A. C. et al. (2012). The effect of isoflurane anaesthesia and vasectomy on circulating corticosterone and ACTH in BALB/c mice. *Gen Comp Endocrinol 179*, 406-413. doi:10.1016/j.ygcen.2012.09.012

Jones, B. C., Sarrieau, A., Reed, C. L. et al. (1998). Contribution of sex and genetics to neuroendocrine adaptation to stress in mice. *Psychoneuroendocrinology 23*, 505-517. doi:10.1016/S0306-4530(98)00014-6

Keubler, L. M., Hoppe, N., Potschka, H. et al. (2020). Where are we heading? Challenges in evidence-based severity assessment. *Lab Anim 54*, 50-62. doi:10.1177/0023677219877216

Kilkenny, C., Parsons, N., Kadyszewski, M. F. et al. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One 4*, e7824. doi:10.1371/journal.pone.0007824

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

Laber, K., Veatch, L. M., Lopez, M. F. et al. (2008). Effects of housing density on weight gain, immune function, behavior, and plasma corticosterone concentrations in BALB/c and C57BL/6 mice. *J Am Assoc Lab Anim Sci 47*, 16-23.

Leenaars, C. H., De Vries, R. B., Heming, A. et al. (2020a). Animal models for cystic fibrosis: A systematic search and mapping review of the literature – Part 1: Genetic models. *Lab Anim 54*, 330-340. doi:10.1177/0023677219868502

Leenaars, C. H., van der Mierden, S., Durst, M. et al. (2020b). Measurement of corticosterone in mice: A protocol for a mapping review. *Lab Anim 54*, 26-32. doi:10.1177/0023677219868499

Lewis, J. and Elder, P. (1985). An enzyme-linked immunosorbent assay (ELISA) for plasma cortisol. *J Steroid Biochem 22*, 673-676. doi:10.1016/0022-4731(85)90222-5

Mulligan, M. K., Ponomarev, I., Boehm, S. et al. (2008). Alcohol trait and transcriptional genomic analysis of C57BL/6 substrains. *Genes Brain Behav 7*, 677-689. doi:10.1111/j.1601-183X.2008.00405.x

Murtagh, R., Behringer, V. and Deschner, T. (2013). LC-MS as a method for non-invasive measurement of steroid hormones and their metabolites in urine and faeces of animals. *Wien Tierärztl Monat – Vet Med Austria 100*, 247-254.

Newsom, S. E. and Darrach, M. (1955). The effect of corticotropin and corticosterone on the production of hemolytic antibodies in the mouse. *Can J Biochem Physiol 33*, 374-379. doi:10.1139/o55-049

Oka, K., Noguchi, M., Kitamura, T. et al. (1987). Liquid chromatography and radioimmunoassay compared for determination of cortisol and corticosterone in plasma after a dexamethasone suppression test. *Clin Chem 33*, 1639-1642. doi:10.1093/clinchem/33.9.1639

Palme, R., Rettenbacher, S., Touma, C. et al. (2005). Stress hormones in mammals and birds: Comparative aspects regarding metabolism, excretion, and noninvasive measurement in fecal samples. *Ann N Y Acad Sci 1040,* 162-171. doi:10.1196/annals.1327.021

Palme, R. (2019). Non-invasive measurement of glucocorticoids: Advances and problems. *Physiol Behav 199*, 299-243. doi:10.1016/j.physbeh.2018.11.021

R Core Team (2019). R: *A Language and Environment for Statistical Computing*. https://www.R-project.org

Ralph, C. and Tilbrook, A. (2016). Invited review: The usefulness of measuring glucocorticoids for assessing animal welfare. *J Anim Sci 94*, 457-470. doi:10.2527/jas.2015-9645

Rød, A. M. K., Harkestad, N., Jellestad, F. K. et al. (2017). Comparison of commercial ELISA assays for quantification of corticosterone in serum. *Sci Rep 7*, 6748. doi:10.1038/s41598-017-06006-4

Russell, W. M. S. and Burch, R. L. (1959). *The Principles of Humane Experimental Technique*. Methuen.

Sheriff, M. J., Dantzer, B., Delehanty, B. et al. (2011). Measuring stress in wildlife: Techniques for quantifying glucocorticoids. *Oecologia 166*, 869-887. doi:10.1007/s00442-011-1943-y

Sittig, L. J., Carbonetto, P., Engel, K. A. et al. (2016). Genetic background limits generalizability of genotype-phenotype relationships. *Neuron 91*, 1253-1259. doi:10.1016/j.neuron.2016.08.013

Smith, A. J., Clutton, R. E., Lilley, E. et al. (2018). PREPARE: Guidelines for planning animal research and testing. *Lab Anim 52*, 135-141. doi:10.1177/0023677217724823

Spackman, D. H. and Riley, V. (1978). Corticosterone concentrations in the mouse. *Science 200*, 87. doi:10.1126/science.635580

Spencer, R. L. and Deak, T. (2017). A users guide to HPA axis research. *Physiol Behav 178*, 43-65. doi:10.1016/j.physbeh.2016.11.014

Spiga, F., Walker, J. J., Terry, J. R. et al. (2011). HPA axis-rhythms. *Compr Physiol 4*, 1273-1298. doi:10.1002/cphy.c140003

Touma, C., Sachser, N., Möstl, E. et al. (2003). Effects of sex and time of day on metabolism and excretion of corticosterone in urine and feces of mice. *Gen Comp Endocrinol 130*, 267-278. doi:10.1016/S0016-6480(02)00620-2

Turpeinen, U. and Hämäläinen, E. (2013). Determination of cortisol in serum, saliva and urine. *Best Pract Res Clin Endocrinol Metab 27*, 795-801. doi:10.1016/j.beem.2013.10.008

Tuttle, A. H., Philip, V. M., Chesler, E. J. et al. (2018). Comparing phenotypic variation between inbred and outbred mice. *Nat Methods 15*, 994-996. doi:10.1038/s41592-018-0224-7

Valentine, H., Williams, W. O. and Maurer, K. J. (2012). Sedation or inhalant anesthesia before euthanasia with CO_2 does not reduce behavioral or physiologic signs of pain and stress in mice. *J Am Assoc Lab Anim Sci 51*, 50-57.

van der Mierden, S., Savelyev, S. A., IntHout, J. et al. (2018). Intracerebral microdialysis of adenosine and adenosine monophosphate – A systematic review and meta-regression analysis of baseline concentrations. *J Neurochem 147*, 58-70. doi:10.1111/jnc.14552

Van Loo, P., Van der Meer, E., Kruitwagen, C. et al. (2004). Long-term effects of husbandry procedures on stress-related parameters in male mice of two strains. *Lab Anim 38*, 169-177. doi:10.1258/002367704322968858

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J Stat Softw 36*, 1-48. doi:10.18637/jss.v036.i03

Windle, R., Wood, S., Shanks, N. et al. (1998). Ultradian rhythm of basal corticosterone release in the female rat: Dynamic interaction with the response to acute stress. *Endocrinology 139*, 443-450. doi:10.1210/endo.139.2.5721

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

The authors would like to thank Alice Tillema for helping with establishing the search strategy, and Rosalie Kempkes and Bobbie Smith for helping with the data extraction. This research was funded by DFG research unit 2591 (BL 953/11-1, TO 542/6-1, JI 276/1-1, LE 2356/5-1, GA427/12-1) and the Federal State of Lower Saxony (R2N).