Behavioral outcomes of complete Freund's adjuvant-induced inflammatory pain in the rodent hind-paw: a systematic review and meta-analysis

Dominika J. Burek[1,2,3], **Nicolas Massaly**[1,3], **Hye Jean Yoon**[1,3], **Michelle Doering**[4], **Jose A. Morón**[1,2,3,5,*]

[1]Department of Anesthesiology, Washington University in St. Louis, St. Louis, MO, 63110, USA;

[2]Department of Neuroscience, Washington University in St. Louis, St. Louis, MO, 63110, USA;

[3]Washington University in St. Louis Pain Center, St. Louis, MO, 63110, USA;

[4]Bernard Becker Medical Library, Washington University School of Medicine, St. Louis, MO, 63110, USA;

[5]Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, 63110, USA

Introduction:

Chronic pain confers higher risk of suicidal ideation and attempt. This public health concern emphasizes the critical need to understand the comorbidity of pain with anxiety and mood disorders [11,26,33]. The shared neural circuitry of nociception and emotional regulation converges these processes for behavioral selection to avoid further injury and promote chances of survival [3]. This integration can become maladaptive in chronic and pathological pain, during which patients may experience heightened stress sensitivity, decreased motivation for routine activity, and anxiety [44,52,55].

Compassionate healthcare aims to provide analgesic relief and to address this emotional component of pain. But progress in drug discovery relies on a strong foundation of preclinical evidence. Innovative therapeutic efforts are stymied by what is lost in translation between human experience and rodent models of subjective states [13,53]. The landscape of preclinical pain research is paved with conflicting evidence, particularly for modeling inflammatory pain and its comorbid syndrome of negative affect [28].

*Correspondence: Jose A. Morón, 660 S. Euclid Avenue, Campus Box 8054, St. Louis, MO 63110, (314) 362-0078, jmoron-concepcion@wustl.edu.

Author contributions:

DJB and JAM designed study;

MD designed and conducted systematic search;

DJB and NM conducted screening and review;

DJB and HJY extracted data;

DJB analyzed data;

DB, NM, and JAM wrote the paper

Injection of Complete Freund's Adjuvant (CFA) into the rodent hind-paw induces sustained inflammation characterized by persistent mechanical hypersensitivity and heat hypersensitivity lasting up to 6 weeks, as our previous work has shown [6,27]. Behavioral alterations in sleep, gait, and food intake have inconsistently been observed but still extensively studied [7,29,50,54]. Other behavioral assessments such as the elevated plus maze (EPM) and the forced swim test (FST), reliable primarily in their predictive validity for antidepressant and anxiolytic drugs, have reasonably been repurposed to expand the available phenotyping of affective changes in pain [6,42,43]. But these assays face limitations in the translational quest to prevent and manage anxiety and mood disorders, and thus have been redefined as modeling aspects of exploratory behavior in aversive environments and coping strategy for inescapable stressors [8,38,60]. Understandably, then, they have provided inconclusive evidence for the effect of neuropathic and persistent pain, including CFA-induced inflammatory pain, on such measures [6,36,43,54,63].

Animal models of pain and their behavioral outcomes must be refined to reproducible designs before further preclinical attempts to identify mechanisms and novel therapeutic avenues. In a valuable reversal of translational direction, systematic reviews and meta-analyses are being adapted from their traditional clinical utility to instead guide preclinical research for this purpose [9]. By highlighting gaps in knowledge, preventing unnecessary replication, and identifying influential variables, preclinical meta-analysis can provide specific strategies for improving the validity and utility of animal models [49,51,56].

The observed heterogeneity in the effect of CFA on measures of exploratory behavior, stress coping, and responses to natural reward may be dependent on heterogeneity of experimental design. The interval between injection and behavioral assessment; the lighting, timing, and duration of phenotyping; and the sex, species, and strain used; the source of experimental animals; and lateralization of paw injection are just some of factors that may influence outcomes [6,30,36,42,43,50,54].

To identify which of these variables influence CFA-induced behavioral alterations and which effects are most often observed, we conducted a meta-analysis. The available relevant literature on the effect of CFA-induced persistent inflammatory pain in rodents, compared to saline exposure, was compiled and analyzed.

Methods:

Protocol development and registration.

A full meta-analysis protocol was reviewed by and registered with the International Prospective Register of Systematic Reviews (PROSPERO CRD42020185408). Methods were informed by resources made available by the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES) consortium and the Systematic Review Centre for Laboratory Animal Experimentation (SYRCLE) [4,49,51,56].

Search.

Search strategies were designed and implemented by a medical librarian (MD) in Ovid Medline 1946-, Embase 1947-, Scopus 1823-, and Web of Science 1900-, and were established using a combination of standardized terms and key words including, but not limited to: (Freunds adjuvant) AND (anxiety OR depression OR emotion OR negative affect OR dark adaptation) AND (mice OR rats). A search in April 2020 found a total of 1,485 results. 903 duplicates were assumed to be accurately identified, bringing the total to 582 unique citations. The literature search was executed again in November 2020. 129 new citations were found. 93 duplicates were removed for a total of 36 unique citations added to the pool of citation results. After importing search results for screening in the systematic review manager Covidence, an additional 6 duplicates were identified, resulting in 612 studies available for screening (Fig. 1A). Search terms are available in Table S1.

Screening.

612 abstracts were imported to Covidence. Two independent reviewers (DJB and NM) conducted abstract screening for studies involving CFA-induced inflammatory pain and behavioral measures in rats and mice. Agreement between reviewers was over 90% and disagreements resolved by consulting a third (JAM).

Inclusion criteria.

Controlled studies with separate treatment (CFA) and control (saline-injected) groups, mice and rats, males and females, adults, all strains, Complete Freund's Adjuvant (AKA Freund's Complete Adjuvant) injection into paws, tests of affective behavior such as the elevated zero or plus maze, open field test (for exploration, not locomotion), forced swim test, tail suspension test, marble-burying, hole-board, social interaction, sucrose preference (AKA sucrose two-bottle choice); tests of aversive behavior such as conditioned place aversion, place escape/avoidance paradigm, mechanical conflict-avoidance assay; operant conditioning such as sucrose progressive ratio or intracranial self-stimulation; wheel running, burrowing assay.

Exclusion criteria.

Studies without a saline-injected control group, randomized clinical trials, within-group controls aka baseline vs. post-CFA injection studies, case studies, in vitro studies, cross-over studies, other animals, neonates, juveniles, models of neuropathic pain such as chronic constriction injury and sciatic nerve injury, models of spinal pain such as spinal nerve ligation, models of acute pain such as formalin or carrageenan injection, CFA injections into body locations other than hind-paw such as knee, tail base, or head, CFA induction of experimental autoimmune encephalomyelitis, studies only evaluating gait or locomotion, naturalistic, home-cage behavior, cognitive & memory assays such as Morris water maze, Barnes maze, novel object recognition.

Full-text review.

113 studies remained after 499 deemed irrelevant. Using Covidence, two independent reviewers (DJB and NM) conducted full-text review with inclusion and exclusion criteria

specified above. Agreement between reviewers was over 90% and disagreements resolved by consulting a third (JAM). 70 studies were excluded: 13 injected CFA into the tail, body, or tibio-tarsal joint, 13 were conference abstracts, 10 studies did not assess affective behavior after an injection of CFA, 8 studies evaluated learning, cognition, or conditioned behavior, 1 evaluated outcomes less than 24 hours after the induction of hindpaw inflammation, and 1 article was not available in English. 4 studies had data in box-and-whisker plots or discrepancies between figures and results, and the authors could not be reached. 4 studies used sham/naïve controls instead of saline injection. 43 studies were ultimately included for systematic review and meta-analysis, plus one additional study published while the analysis was ongoing bringing the total to 44 included studies (Fig. 1A).

Quality assessments of risks of bias.

The SYRCLE risks-of-bias tool was adapted and customized with criteria available in Table S2. If a criterion was not mentioned or reported, or was unclear, it was coded as unknown rather than assumed missing for high risk of bias. Using Covidence, two independent researchers (DB and NM) evaluated each study. Agreement between reviewers was over 90% and disagreements resolved by consulting a third (JAM).

Data extraction and abstraction.

Two reviewers (DB and HJY) worked independently and in parallel. The following study and experimental details were extracted to a Microsoft Excel spreadsheet: bibliographical: first author, last author, year published, journal; animal variables: species, source, light cycle, age or weight, strain, sex, housing; CFA methodology: number of injections/ exposures, hind-paw side injected, volume, concentration, manufacturer, validation method for heat hypersensitivity or mechanical allodynia, interval between injection and behavioral assessment; behavioral test-specific methodology: lighting, duration, whether animal was naïve to other behavioral testing, whether baselines were conducted. Time spent in the open arms for EPM/EZM, time spent in the center for OFT, time spent in the light compartment for LDB, time spent in the dark/stimulation compartment for PEAP, time spent immobile for FST and TST, percent preference for SPT, weight of substrate burrowed, and distance or revolutions for wheel running were extracted. Study authors were reached through email for unreported variables. Variables that were not explicitly reported but instead cited another paper from the same authors (for example, "EPM was conducted as previously described" and lighting or duration were omitted in the study in question) were tracked to the original cited paper. Some information was still unavailable; experiments with unknown variables were excluded from relevant but not all subgroup analyses. No study reported means and standard error of the mean in the results; outcome data represented graphically were abstracted using the digital ruler WebPlot Digitizer.

Data analysis.

Individual, whole, published articles are referred to as "studies;" discrete cohorts within articles that differ by variables such as sex are referred to as "experiments;" and experiments with repeated measures data are referred to as such (Fig. 1B). Discrete meta-analyses were conducted for 9 behavioral tests separately: elevated plus or zero maze (EPM/EZM), open field test (OFT), light-dark box (LDB), forced swim test (FST), tail suspension test (TST),

place escape/avoidance paradigm (PEAP), sucrose preference test (SPT), wheel running (WHEEL), and burrowing (BURR) (Fig. 1C). Data analysis was conducted using the *meta*, *metafor*, and *dmetar* R packages [19,57]. Standardized mean difference effect sizes were calculated to Hedge's *g* with 95% confidence intervals. Experiments were summarized in a random-effects meta-analysis, using the restricted maximum-likelihood estimation of heterogeneity with the Hartung and Knapp correction, and the inverse variance method of weighing [4,56]. Summary effects statistics include 95% confidence intervals (range where summary effect mean lies within 95% certainty) and 95% prediction intervals (predicted range for 95% of future effects based on heterogeneity). These analyses were followed by sub-group AKA stratified meta-analysis for categorical variables and meta-regression for continuous variables [4,56]. Sub-group identities overlapped with each other. Heterogeneity was estimated using tau (τ) and Higgin's and Thompson's I^2 in global and sub-group/ stratified meta-analysis; Q tests for residual heterogeneity and moderators and R^2 for meta-regression [4,19,20,56,57]. A Bonferroni correction for multiple comparisons adjusted the threshold for significance to $p<0.007$ for 7 total sub-group analyses meta-regressions. For publication bias, separate analyses were also conducted by nesting experiments from the same study into a fixed-effects model summary, as individual experiments are not published but rather grouped into studies and published as a whole. Whole studies were then analyzed in a random-effects model of a global summary. These analyses were followed by funnel plots, Egger's regression, and trim-and-fill analysis. Figures were created using the *ggplot2* R package [61].

Open-source data.

Data, results, and code are provided on the Open Science Framework https://osf.io/tzbje/? view_only=55051ed2845d4f28a9602122889defe4. Individual spreadsheets for all 9 of the behavioral tests listed above provide bibliographical, methodological, and quantitative data for each experiment; meta-analysis results by experiment and by study; and sub-group analysis and meta-regression results. Individual effect sizes and confidence intervals for studies which used sham instead of saline-injected controls are provided in an "Exclusions" spreadsheet, in addition to a random-effects summary estimate if they had been included in the relevant meta-analysis. Individual effect sizes and confidence intervals for behavioral paradigms appearing in only one study (thermal operant, intra-cranial self-stimulation, nesting, hole-board, marble-burying, splash test, progressive ratio, and social interaction) are provided in a "Miscellaneous" spreadsheet. R script including code for conducting analyses, exporting results, and generating plots is also provided.

Data presentation: forest plots.

Effect sizes are plotted as circles with 95% confidence intervals marked by the underlying line. Size of circle represents its weight in the random-effects summary. Color of circle matches a symmetrical scale extending from the lowest effect size in darkest purple to the highest effect size in green, with zero as white. Color scale legend is overlayed on the x-axis and is tailored to the range of Hedge's *g* for each individual meta-analysis per behavioral test. Dotted line provides a marker for x=0. Bottom diamond represents the random-effects summary effect size and 95% confidence intervals. The underlying cropped color scale represents the prediction interval. Dashed line represents the marker for the summary effect.

Data presentation: bubble plots.

Effect size is on the y-axis and levels of the continuous variable moderator are on the x-axis. Color and size scales are the same as the experiment forest plot.

Data presentation: funnel plots.

Effect size is on the x-axis and 1/ N is on the y-axis. Color scale is the same as the experiment forest plot. Gray circles are theoretical studies added by trim-and-fill analysis.

Results:

Assessment of study quality and risks of bias.

100% of studies reported compliance with standards of animal welfare or approval of animal care and use by institutional committee (Fig. 2). 88.6% of studies did not report any method of determining experimental sample sizes. 2 studies reported conducting a power analysis, and 3 studies reported taking into consideration the minimal number needed based on previous data. 90.9% of studies did not report randomized order of testing. 4 studies reported random selection or counterbalancing but did not specify exact method. 81.8% of studies reported no private funding or affiliation. 2 studies did not include a funding disclosure, 5 studies reported corporate funding, and 1 study reported no private funding for the study in question, but authors were affiliated with and compensated by private entities. Similarly, 79.6% of studies reported no conflicts of or competing interests; however, 6 studies did not include a statement whatsoever and 3 studies reported authors with patents or corporate affiliation. 63.6% of studies did not report randomizing allocation of saline vs. CFA injection. 16 studies reported randomly choosing animals for injections but did not specify exact method. Since the hind-paw inflammation induced by CFA is visually obvious, it precludes conducting experiments blind; however, of studies with variables other than saline vs. CFA, 48.7% reported blinding experimenters and 3 studies explicitly stated no blinding. All studies reported similar baseline population characteristics such as age and weight-matching experimental groups, but most relevant for studies featuring a model of inflammatory pain was the confirmation of the mechanical allodynia and/or heat hypersensitivity induced by CFA injection. 54.6% of studies conducted von Frey or Hargreaves assays on the same animals evaluated in affective behavioral measures. 11 studies did not conduct any assessment. 63.6% of studies did not report exclusions or absence thereof. 5 studies confirmed no exclusions and 11 studies cited exclusion criteria based on baseline performance. 47.7% of studies selectively reported outcomes by omitting locomotor activity measures or secondary measures such as entries into the open arms of the elevated maze, or light compartment of the light-dark box. 65.9% of studies reported fully automated and/or blinded manual behavioral scoring and analysis. 3 studies reported unblinded manual scoring (Fig. 2).

Summary effect sizes for 9 behavioral tests.

Forest plots by experiment, by study, using the first of repeated measures, and using the greatest effect in repeated measures are available for all 9 of these behavioral tests in the supplemental figures cited throughout the following results.

CFA significantly reduced time spent in the open arms of elevated plus and zero mazes (g=−0.5881 [Confidence Interval (CI) −0.8738 to −0.3024; Prediction Interval (PI) −1.8413 to 0.6651], p=0.0002, I^2=54.6%, τ=0.6010, k=37, N=580) (Fig. 3, S1). 70.3% of experiments included had individually non-significant effects. Heterogeneity was moderate. The prediction interval suggests most future studies will not indicate a robust and significant effect of CFA on exploratory behavior in an EPM/EZM.

CFA significantly reduced time spent in the center of an open field arena (g=−0.2910 [CI −0.5682 to −0.0138; PI −1.4851 to 0.9031], p=0.0401, I^2=56.2%, τ=0.5742, k=41, N=679; Fig. 3, S2). 80.5% of experiments included had individually non-significant effects. Heterogeneity was moderate. The prediction interval suggests most future studies will not indicate a robust and significant effect of CFA on exploratory behavior in an OFT.

CFA significantly reduced time spent in the light compartment of a light-dark box (g=−0.6369 [CI −1.1137 to −0.1602; PI −1.8608 to 0.5870], p=0.0118, I^2=47.6%, τ=0.5313, k=18, N=210; Fig. 3, S3). 83.3% of experiments included had individually non-significant effects. Heterogeneity was moderate. The prediction interval suggests most future studies will not indicate a robust and significant effect of CFA on exploratory behavior in an LDB.

CFA significantly reduced time spent in a dark compartment with noxious hind-paw stimulation in the place escape/avoidance paradigm (g=−2.2156 [CI −3.4905 to −0.9408; PI −5.9046 to 1.4733], p=0.0039, I^2=82.0%, τ=1.4588, k=9, N=198; Fig. 3, S4). Only 1 experiment included had a non-significant effect. Heterogeneity was very high. The prediction interval suggests most future studies could indicate CFA decreases the time spent in the aversive environment in the PEAP.

CFA did not significantly increase immobility in the forced swim test (g=0.1527 [CI −0.2749 to 0.5802; PI −1.0140 to 1.3193], p=0.4601, I^2=50.8%, τ=0.5088, k=17, N=280; Fig. 3, S5). 82.4% of experiments included had individually non-significant effects. Heterogeneity was moderate. The prediction interval suggests most future studies will also not indicate a robust and significant effect of CFA on immobility in the FST.

CFA did significantly increase immobility in the tail suspension test (g=2.7387 [CI 1.6723 to 3.8050; PI −0.4949 to 5.9722], p=0.0001, I^2=71.4%, τ=1.3680, k=12, N=152; Fig. 3, S6. Only 16.7% of experiments included had individually non-significant effects; however, 83.3% of experiments included came from a single study. Heterogeneity was very high. The prediction interval suggests most future studies could still indicate CFA increases immobility in the TST.

CFA significantly reduced sucrose preference in the first trial at which animals were naïve to repeated measures (g=−0.4489 [CI −0.8936 to −0.0041; PI −0.9292 to 0.0315], p=0.0486, I^2=0.0%, τ=0.0008, k=6, N=140;. Fig 3, S7–8). 66.7% of experiments included had individually non-significant effects. Heterogeneity was extremely low. The prediction interval suggests most future studies may not indicate a robust and significant effect of CFA on sucrose preference.

CFA significantly reduced revolutions or distance traveled while wheel running during a first trial at which animals were naïve to repeated measures ($g=-2.2074$ [CI -2.6016 to -1.8132; PI -2.6044 to -1.8103], $p<0.0001$, $I^2=0.0\%$, $\tau=0.0000$, $k=15$, $N=213$; Fig. 3, S9–10). Only 13.3% of experiments included had individually non-significant effects. Heterogeneity was very low. The prediction interval suggests most future studies could further indicate CFA decreases wheel running within a narrow range of possible effect in this paradigm.

CFA significantly reduced mass of substrate burrowed during a first trial at which animals were naïve to repeated measures ($g=-2.2323$ [CI -2.9460 to -1.5186; PI -4.6967 to 0.232], $p<0.0001$, $I^2=73.9\%$, $\tau=1.0992$, $k=16$, $N=299$; Fig. 3, S11–12). 18.8% of experiments included had individually non-significant effects. Heterogeneity was very high. The prediction interval suggests most future studies will also indicate a significant decrease in burrowing by CFA.

Sub-group analysis and meta-regression.

Sub-group analysis and meta-regression were conducted to identify possible sources of heterogeneity. The burrowing assay, elevated plus/zero maze, and forced swim test were primary paradigms of interest given the large effect of CFA on burrowing, and the particularly common use of the EPM/EZM and FST in behavioral phenotyping. Animal source, strain, sex, and paw injected were primary categorical variables of interest given their particular variation in experimental methods. Sub-group forest plots for additional categorical variables such as C57 sub-strain and lighting conditions for all 9 of these behavioral tests are provided in the supplemental figures cited throughout the following results.

Sub-group analysis of burrowing experiments revealed animal sourcing to account for a significant proportion of heterogeneity ($Q_2=55.81$, $p<0.0001$) such that animals purchased from Charles River exhibited the greatest burrowing deficit, although only one and two experiments sourced animals from an institutional colony or Jackson Laboratories, respectively (Fig. 4A, S13–14). Strain also accounted for a significant proportion of heterogeneity in burrowing ($Q_2=55.81$, $p<0.0001$) although this was primarily due to group membership overlap in which all Wistar-Han rats were purchased from Charles River. Only one experiment used female rodents, which experienced no significant effect of CFA on their burrowing behavior, leading to sex being a significant contributor to heterogeneity ($Q_1=11.36$, $p=0.0008$). Injection laterality accounted for a significant proportion of heterogeneity, with bilaterally injected animals exhibiting the greatest burrowing deficits ($Q_2=67.55$, $p<0.0001$) (Fig. 4A, S13–14). Extended subgroup analysis is included in Fig. S13A, 14A.

Sub-group analysis of EPM and EZM experiments also indicated significant influence of animal sourcing on CFA-reduced exploratory behavior, such that CFA-injected animals raised in institutional colonies rather than purchased from a vendor explored elevated mazes the least ($Q_3=15.61$, $p=0.0014$) (4B, S15). Strain did not significantly contribute to heterogeneity in reduction in exploratory behavior by CFA ($Q_3=11.80$, $p=0.0081$ (Bonferroni correction). Although almost all the animals used were of the C57BL/6 background, the effect of CFA was greater on Sprague-Dawley rats and negligible on

FVB/NJNju and Balb/c mice. Neither sex nor paw laterality significantly accounted for heterogeneity (Q_1=0.42, p=0.5194; Q_1=1.4656, p=0.2261) (Fig. 4B, S15). Extended subgroup analysis is included in Fig. S15A.

Immobility in the FST was not significantly affected by animal sourcing, although the greatest increase in immobility was observed in the one experiment using rodents sourced from Taconic, with no significant effect of CFA in experiments using Charles River (Q_3=9.17, p=0.0272 (Bonferroni correction)) (Fig. 4C, S16). Strain did not account for a significant proportion of heterogeneity in CFA-induced forced swim immobility, although the majority of experiments used C57BL/6 mice which overall had the lowest magnitude of effect (Q_4=12.14, p=0.0164 (Bonferroni correction)). Neither sex nor paw laterality significantly accounted for heterogeneity (Q_1=0.07, p=0.7972; Q_1=0.03, p=0.8751) (Fig. 4C, S16). Extended subgroup analysis is included in Fig. S16A.

Sample size in the control group, time point tested post-CFA injection, and micrograms of CFA injected were primary continuous variables of interest given their particular variation in experimental methods. Bubble plots for additional continuous variable such as age and duration of assessment for all 9 of these behavioral tests are provided in the supplemental figures cited throughout the following results.

Sample size in the control, saline-injected group was not significantly associated with effect size and accounted for only 5.4% of observed heterogeneity in burrowing deficits ($F_{1,14}$=1.14, p=0.3029) (Fig. 4D, 13–14). Because the initial trial of any repeated measures was chosen to conduct this meta-analysis, no significant differences were observed between animals tested on day 1 or 2 post-CFA injection, but the data are provided for comparison regardless ($F_{1,14}$=0.20, p=0.6581). The amount of CFA, in micrograms, injected into the paw to induce inflammation was not a moderator of effect, accounting for 39.72% of observed heterogeneity albeit with a significant amount of heterogeneity left over ($F_{1,14}$=6.06, p=0.0274 (Bonferroni correction)) (Fig. 4D). Additional meta-regressions, including for analyses conducted on data point with greatest effect from repeated measures, are provided in Fig. S13B, 14B.

Control-group sample size accounted for 0.0% of the observed heterogeneity in CFA-induced exploratory behavior deficits in an EPM or EZM ($F_{1,35}$=0.01, p=0.9231) (Fig. 4E). Although animals included in this meta-analysis were tested anywhere from one to 42 days post-CFA-injection, interval to assessment accounted for 0.0% of observed heterogeneity ($F_{1,35}$=0.17, p=0.6831). Amount of CFA injected was not a significant moderator of effect, accounting for only 1.8% of observed heterogeneity ($F_{1,35}$=1.26, p=0.2694) (Fig. 4E). Additional meta-regressions are included in Fig. S13B, 14B.

Amount of CFA injected accounted for 18.9% of heterogeneity in FST experiments but was not a significant predictor of treatment effect ($F_{1,15}$=2.16, p=0.1620) (Fig. 4F). Sample size in the control group accounted for 19.1% of heterogeneity, but was not significant ($F_{1,15}$=1.74, p=0.2067) (Fig. 4F). Interval to assessment, in days, accounted for 9.2% of heterogeneity but was not a significant moderator of effect either ($F_{1,15}$=1.52, p=0.2365). Additional meta-regressions are included in Fig. S16B.

All sub-group analyses and meta-regressions for OFT, LDB, PEAP, TST, SPT, and wheel running are included in Fig. S17–S24. Summary tables for significance are available in Tables S4 and S5.

Publication bias.

To determine publication bias, experiments from the same study were nested to analyze whole studies, which are evaluated as a unit for publication unlike individual experiments. Experiments within studies were nested with a fixed-effects model to account for being conducted in the same environment, with the same resources, by the same people. Whole studies were then analyzed in a random-effects model, then represented in funnel plots, tested for asymmetry with Egger's regression, and analyzed with trim-and-fill. Publication bias analyses are most reliable with greater than 10 studies, which were only available for the EPM/EZM and OFT. Analyses are provided for the other 7 behavioral paradigms regardless for comparison.

There was still a significant effect of CFA on exploratory behavior in the EPM/EZM (by study −0.8982 [CI −1.2740 TO −0.5223, PI −2.2755 to 0.4792], p<0.0001, τ=0.6171, I^2=72.3%, k=17, N=580) and Egger's regression indicated the presence of funnel plot asymmetry (intercept −3.135, p=0.0002) (Fig. 5A). Trim-and-fill analysis added 8 studies with either an increase or no change in open arm time (g=−0.3918 [CI −0.7871 TO 0.0034], p=0.0520) (Fig. 5A).

There was still a significant effect of CFA on exploratory behavior in the OFT (by study −0.5237 [CI −0.8837 to −0.1636, PI −1.8623 to 0.8150], p=0.0044, τ=0.6042, I^2=73.8%, k=18, N=679) and Egger's regression indicated the presence of funnel plot asymmetry (intercept −2.129, p=0.0145) (Fig. 5B). Trim-and-fill analysis added 5 studies with either an increase or no change in center time (g=−0.1463 [CI −0.5398 to 0.2473], p=0.4663) (Fig. 5B).

Trim-and-fill analysis and Egger's regression lack the statistical power to reliably and accurately detect bias with fewer than 10 studies. Funnel plots of "by experiment" effect sizes vs. 1/ N, *without* trim-and-filled studies, for LDB, PEAP, FST, TST, SPT, wheel running, and burrowing are available in Fig. S25.

Discussion:

This systematic review and meta-analysis identified significant effects of CFA, and by extension, its induced persistent inflammatory pain, on several classical tests of exploratory behavior, stress coping, and naturalistic behavior. With data from dozens of experiments and hundreds of mice and rats, we show that CFA significantly decreases exploratory behavior by reducing open arm time in the EPM/EZM, center time in the OFT, and light compartment time in the LDB. CFA also significantly increases immobility in the TST but not the FST. The magnitude of effect was often extremely modest, and prediction intervals inconclusively wide, compared to those identified in other preclinical meta-analyses [9]. The greatest negative effects can be observed in naturalistic behaviors such as burrowing and wheel running. Subgroup analyses most often revealed animal sourcing from an institutional

colony as associated with a greater magnitude of CFA effect. Significant differences were also frequently found between species and strains.

The conclusions of these subgroup analyses and meta-regressions were constrained by subgroup membership sample size. The validity of these statistical methods relies on greater than 10 comparisons; otherwise, significance appears spurious. In this meta-analysis, although 40 experiments were conducted using the OFT, only one evaluated Balb/c mice and two assessed rats. The significance of species and strain differences is thus more of a clue towards a prediction interval for Balb/c mice and rats, rather than a confirmation of what drives true and observed CFA effects.

Regardless, animal sourcing was identified as accounting for a significant proportion of heterogeneity in EPM, OFT, LDB, FST, wheel running, and burrowing experiments. The differences between research animal vendors and institutional colonies are still subject to speculation, although factors such as genetic drift and shipping conditions may be confounding the effects of CFA on these behavioral outcomes. Moreover, we expected that lateralization of CFA injection would be a significant contributor to heterogeneity. Most studies included here injected CFA unilaterally, with a roughly equal split between right vs. left paws. "Paws" were a significant contributor to heterogeneity in burrowing experiments, such that bilaterally injected animals exhibited the greatest CFA-induced deficits, and to a modest degree in sucrose preference as well. The same was not observed in wheel running or other paradigms, despite previous studies describing differences in exercise, activity, locomotion, and gait between unilaterally- and bilaterally-injected animals [7,50].

Another limitation of this work, and of many preclinical meta-analyses by nature, is the reliance on accurate and detailed methods reporting. Group vs. single-housing conditions, environmental enrichment if any, handling and habituation protocols, and bedding presence or absence in the testing apparatus are an incomplete list of factors that may influence animal behavior but are rarely reported [5,14,17,46,59]. Not only is reproducibility impeded by methodological opacity, but also the strategy for novel experiments which this meta-analysis aims to provide.

To some degree, this may be explained by evolving journal and field standards; for example, the Animal Research: Reporting in Vivo Experiments (ARRIVE) guidelines were not published until 2010 [26]. This checklist for standardized reporting of methods, however, is far from standardized methods themselves. Researchers frequently customize and modify protocols. Affective phenotyping already suffers from very limited and hotly debated face and construct validity [18,38]. Tenuous to begin with, these are further eroded when lighting, duration of assessment, and height of maze or depth of water are independent variables themselves [2,17,35,58]. Moreover, the environment and paradigm detect the same symptoms they evoke [13]. The experience is itself anxiogenic and energy-depleting.

The predictive validity of the EPM (and the OFT, LDB, and "tests of unconditioned anxiety" overall) and of the FST as it applies to persistent or chronic pain is also doubtful, while also being the primary appeal in leveraging these models [8,12,13,43]. Behavioral deficits in exploratory behavior are typically corrected by the anxiolytic benzodiazepine

drugs, which augment the baseline of wild-type and naïve animals regardless [17,35]. Benzodiazepines are all but irrelevant to the CFA model of inflammatory pain, as current clinical guidelines advise against their prescription to patients with inflammatory conditions [39,62]. Similarly, immobility in the FST and TST is sensitive to antidepressants, primarily selective serotonin reuptake inhibitors, even in wild-type and naïve mice [25]. The analgesia conferred by SSRIs is inconsistent in humans and recommended primarily for acute or mixed-chronic pain, but again not for inflammatory conditions such as fibromyalgia [24]. SSRIs are ineffective on neuropathic pain too, eclipsed only marginally by tricyclics [47]. (Pharmacological interventions and other manipulations from the studies included in this meta-analysis are summarized in Table S6.)

Of course, it is imperative to equally question the validity of CFA itself, and to what degree that is a model of long-term inflammatory pain. Our lab has previously demonstrated that CFA-induced heat hypersensitivity, mechanical allodynia, and paw swelling can last for six weeks in male C57BL/6J mice, so it is certainly a persistent condition with enough time to develop other symptoms, compared to other behavior-altering courses such as chronic unpredictable/mild stress [6,32].

Furthermore, inflammation and mood have a well-established, if poorly understood, reciprocal relationship: for example, cancer patients given recombinant cytokines suffer from reduced appetite, sleeplessness, social withdrawal, and lethargy [11,45]. In one theory of chronic pain, behavior is modified to avoid further injury and ensure convalescence, but in the absence of physical healing and sustained inflammation, this maladaptive process turns pathological [3]. In rodents, this may appear more as motivational deficits for natural rewards such as sucrose pellets, burrowing opportunities, nestlet building, social interaction, wheel running, and pain relief [7,16,21,33,37,41,42,48]. Motivational deficits may also be accompanied by alterations in hedonic state and processing, such as reduced sucrose intake without a decrease in preference, as our own recent and emerging data suggest [21,33]. Reduced participation in or motivation for such activities by rodents could be viewed as comparable to human patients avoiding physical therapy, neglecting their hobbies or home, experiencing anhedonia, and socially withdrawing.

These paradigms are also closer to the Research Domain Criteria's priorities on individual behaviors and individual symptoms on a continuum of typical to pathological [1,10,23]. Stratifying and classifying clusters of symptoms was always intended for diagnostic purposes, not because evidence has always endorsed common pathogenesis [1]. In all these attempts at modeling the physical, cognitive, emotional, and sensory components in a syndrome of chronic pain, the common pathogenesis is still undefined, and thus fails to justify continued use of the battery of tests this meta-analysis evaluates. Faced with such unconvincing data that precludes answering mechanistic questions, preclinical and basic pain researchers could turn to evaluating naturalistic behavior.

In rodents, decreased motivation for sucrose pellets in a progressive ratio task, disturbances in slow-wave sleep, reduced social interaction, and decreased wheel running and burrowing, have already identified some neural correlates and some mechanisms of inflammatory pain's effect on the brain [7,16,29,33,41,42,48]. Although the number of experiments and

studies conducted so far is small and methodologically heterogeneous, they have been successfully replicated. These paradigms do have their own limitations, including possible locomotor confounds, requirements of food or water deprivation, and testing outside the home-cage. They are also, however, compatible with repeated measures, automation, and deep phenotyping with sophisticated and agnostic tools. Open-source programs such as Deep Lab Cut, Simple Behavioral Analysis, and Bonsai are accessible and customizable, allowing for unbiased approaches to evaluating nuanced and subtle behaviors both healthy and pathological [31,34,40].

This is a hopeful and promising alternative path forward. Many of the classical phenotyping assays featured in this work indicated significant effects of CFA in a random-effects summary estimate, but the magnitude of effect was modest compared to those identified in other preclinical meta-analyses. The body of literature analyzed here intends to be a strong foundation of basic research that builds to human clinical trials, but the heterogeneity in both methods and outcomes is the rate-limiting step for progress in chronic pain therapies.

While previous efforts have provided thorough and informative narrative reviews, this study quantified the effects of CFA on nine commonly used measures of pain-reduced behaviors and stratified these effects by variables of interest such as age, sex, species, animal source, and many others [15,28]. Future research can be refined and optimized when so informed on the choice of behavioral paradigm, population, and even sample size. These considerations while leveraging motivated and naturalistic behavior, next-generation animal-tracking, and the creative development of novel behavioral paradigms, will bring us closer and closer to helping the millions of people living with chronic pain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References:

[1]. Anderzhanova E, Kirmeier T, Wotjak CT. Animal models in psychiatric research: The RDoC system as a new framework for endophenotype-oriented translational neuroscience. Neurobiol Stress 2017;7:47–56. [PubMed: 28377991]

[2]. Arabo A, Potier C, Ollivier G, Lorivel T, Roy V. Temporal analysis of free exploration of an elevated plus-maze in mice. J Exp Psychol Anim Learn Cogn 2014;40:457–466. [PubMed: 25546103]

[3]. Baliki MN, Apkarian AV. Nociception, Pain, Negative Moods, and Behavior Selection. Neuron 2015;87:474–491. [PubMed: 26247858]

[4]. Borenstein M ed. Introduction to meta-analysis. Chichester, U.K: John Wiley & Sons, 2009.

[5]. Bravo L, Alba-Delgado C, Torres-Sanchez S, Mico JA, Neto FL, Berrocoso E. Social stress exacerbates the aversion to painful experiences in rats exposed to chronic pain: The role of the locus coeruleus. PAIN® 2013;154:2014–2023. [PubMed: 23792242]

[6]. Burek DJ, Massaly N, Doering M, Zec A, Gaelen J, Morón JA. Long-term inflammatory pain does not impact exploratory behavior and stress coping strategies in mice. PAIN 2021;Articles in Press. doi:10.1097/j.pain.0000000000002179.

[7]. Cobos EJ, Ghasemlou N, Araldi D, Segal D, Duong K, Woolf CJ. Inflammation-induced decrease in voluntary wheel running in mice: a nonreflexive test for evaluating inflammatory pain and analgesia. Pain 2012;153:876–884. [PubMed: 22341563]

[8]. Commons KG, Cholanians AB, Babb JA, Ehlinger DG. The Rodent Forced Swim Test Measures Stress-Coping Strategy, Not Depression-like Behavior. ACS Chem Neurosci 2017;8:955–960. [PubMed: 28287253]

[9]. Currie GL, Angel-Scott HN, Colvin L, Cramond F, Hair K, Khandoker L, Liao J, Macleod M, McCann SK, Morland R, Sherratt N, Stewart R, Tanriver-Ayder E, Thomas J, Wang Q, Wodarski R, Xiong R, Rice ASC, Sena ES. Animal models of chemotherapy-induced peripheral neuropathy: A machine-assisted systematic review and meta-analysis. PLOS Biology 2019;17:e3000243. [PubMed: 31107871]

[10]. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med 2013;11:126. [PubMed: 23672542]

[11]. Dantzer R, Kelley KW. Twenty years of research on cytokine-induced sickness behavior. Brain Behav Immun 2007;21:153–160. [PubMed: 17088043]

[12]. Ennaceur A Tests of unconditioned anxiety — Pitfalls and disappointments. Physiology & Behavior 2014;135:55–71. [PubMed: 24910138]

[13]. Ennaceur A, Chazot PL. Preclinical animal anxiety research – flaws and prejudices. Pharmacol Res Perspect 2016;4. doi:10.1002/prp2.223.

[14]. Gabriel AF, Paoletti G, Seta DD, Panelli R, Marcus MAE, Farabollini F, Carli G, Joosten EAJ. Enriched environment and the recovery from inflammatory pain: Social versus physical aspects and their interaction. Behavioural Brain Research 2010;208:90–95. [PubMed: 19914294]

[15]. González-Cano R, Montilla-García Á, Ruiz-Cantero MC, Bravo-Caparrós I, Tejada MÁ, Nieto FR, Cobos EJ. The search for translational pain outcomes to refine analgesic development: Where did we come from and where are we going? Neuroscience & Biobehavioral Reviews 2020;113:238–261. [PubMed: 32147529]

[16]. Gould SA, Doods H, Lamla T, Pekcec A. Pharmacological characterization of intraplantar Complete Freund's Adjuvant-induced burrowing deficits. Behavioural Brain Research 2016;301:142–151. [PubMed: 26704218]

[17]. Gould TD ed. Mood and Anxiety Related Phenotypes in Mice: Characterization Using Behavioral Tests. Humana Press, 2009 doi:10.1007/978-1-60761-303-9.

[18]. Gururajan A, Reif A, Cryan JF, Slattery DA. The future of rodent models in depression research. Nature Reviews Neuroscience 2019;20:686–701. [PubMed: 31578460]

[19]. Harrer M, Cuijpers P, Furukawa T, Ebert D. Doing Meta-Analysis in R: A Hands-on Guide. 2019.

[20]. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine 2002;21:1539–1558. [PubMed: 12111919]

[21]. Hipólito L, Wilson-Poe A, Campos-Jurado Y, Zhong E, Gonzalez-Romero J, Virag L, Whittington R, Comer SD, Carlton SM, Walker BM, Bruchas MR, Morón JA. Inflammatory Pain Promotes Increased Opioid Self-Administration: Role of Dysregulated Ventral Tegmental Area μ Opioid Receptors. J Neurosci 2015;35:12217–12231. [PubMed: 26338332]

[22]. Hooten WM. Chronic Pain and Mental Health Disorders: Shared Neural Mechanisms, Epidemiology, and Treatment. Mayo Clin Proc 2016;91:955–970. [PubMed: 27344405]

[23]. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010;167:748–751. [PubMed: 20595427]

[24]. Jung AC, Staiger T, Sullivan M. The Efficacy of Selective Serotonin Reuptake Inhibitors for the Management of Chronic Pain. 1997;12:6.

[25]. Kara NZ, Stukalin Y, Einat H. Revisiting the validity of the mouse forced swim test: Systematic review and meta-analysis of the effects of prototypic antidepressants. Neuroscience & Biobehavioral Reviews 2018;84:1–11. [PubMed: 29128579]

[26]. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLOS Biology 2010;8:e1000412. [PubMed: 20613859]

[27]. Knight B, Katz DR, Isenberg DA, Ibrahim MA, Le Page S, Hutchings P, Schwartz RS, Cooke A. Induction of adjuvant arthritis in mice. Clin Exp Immunol 1992;90:459–465. [PubMed: 1458683]

[28]. Kremer M, Becker LJ, Barrot M, Yalcin I. How to study anxiety and depression in rodent models of chronic pain? European Journal of Neuroscience 2021;53:236–270.

[29]. Leys LJ, Chu KL, Xu J, Pai M, Yang HS, Robb HM, Jarvis MF, Radek RJ, McGaraughty S. Disturbances in slow-wave sleep are induced by models of bilateral inflammation, neuropathic, and postoperative pain, but not osteoarthritic pain in rats. Pain 2013;154:1092–1102. [PubMed: 23664655]

[30]. Liu Y, Yang L, Yu J, Zhang Y-Q. Persistent, comorbid pain and anxiety can be uncoupled in a mouse model. Physiol Behav 2015;151:55–63. [PubMed: 26166157]

[31]. Lopes G, Bonacchi N, Frazão J, Neto JP, Atallah BV, Soares S, Moreira L, Matias S, Itskov PM, Correia PA, Medina RE, Calcaterra L, Dreosti E, Paton JJ, Kampff AR. Bonsai: an event-based framework for processing and controlling data streams. Front Neuroinform 2015;9. doi:10.3389/fninf.2015.00007.

[32]. Lu Q, Mouri A, Yang Y, Kunisawa K, Teshigawara T, Hirakawa M, Mori Y, Yamamoto Y, Libo Z, Nabeshima T, Saito K. Chronic unpredictable mild stress-induced behavioral changes are coupled with dopaminergic hyperfunction and serotonergic hypofunction in mouse models of depression. Behav Brain Res 2019;372:112053. [PubMed: 31288060]

[33]. Massaly N, Copits BA, Wilson-Poe AR, Hipólito L, Markovic T, Yoon HJ, Liu S, Walicki MC, Bhatti DL, Sirohi S, Klaas A, Walker BM, Neve R, Cahill CM, Shoghi KI, Gereau RW, McCall JG, Al-Hasani R, Bruchas MR, Morón JA. Pain-Induced Negative Affect Is Mediated via Recruitment of The Nucleus Accumbens Kappa Opioid System. Neuron 2019;102:564–573.e6. [PubMed: 30878290]

[34]. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature Neuroscience 2018;21:1281–1289. [PubMed: 30127430]

[35]. Mechan AO, Moran PM, Elliott MJ, Young AM, Joseph MH, Green RA. A comparison between Dark Agouti and Sprague-Dawley rats in their behaviour on the elevated plus-maze, open-field apparatus and activity meters, and their response to diazepam. Psychopharmacology 2002;159:188–195. [PubMed: 11862348]

[36]. Narita M, Kaneko C, Miyoshi K, Nagumo Y, Kuzumaki N, Nakajima M, Nanjo K, Matsuzawa K, Yamazaki M, Suzuki T. Chronic pain induces anxiety with concomitant changes in opioidergic function in the amygdala. Neuropsychopharmacology 2006;31:739–750. [PubMed: 16123756]

[37]. Negus SS, Neddenriep B, Altarifi AA, Carroll FI, Leitl MD, Miller LL. Effects of ketoprofen, morphine, and kappa opioids on pain-related depression of nesting in mice. Pain 2015;156:1153–1160. [PubMed: 25827812]

[38]. Nestler EJ, Hyman SE. Animal models of neuropsychiatric disorders. Nature Neuroscience 2010;13:1161–1169. [PubMed: 20877280]

[39]. Nielsen S, Lintzeris N, Bruno R, Campbell G, Larance B, Hall W, Hoban B, Cohen ML, Degenhardt L. Benzodiazepine Use among Chronic Pain Patients Prescribed Opioids: Associations with Pain, Physical and Mental Health, and Health Service Utilization. Pain Med 2015;16:356–366. [PubMed: 25279706]

[40]. Nilsson SR, Goodwin NL, Choong JJ, Hwang S, Wright HR, Norville ZC, Tong X, Lin D, Bentzley BS, Eshel N, McLaughlin RJ, Golden SA. Simple Behavioral Analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals. bioRxiv 2020:2020.04.19.049452.

[41]. Okun A, McKinzie DL, Witkin JM, Remeniuk B, Husein O, Gleason SD, Oyarzo J, Navratilova E, McElroy B, Cowen S, Kennedy JD, Porreca F. Hedonic and motivational responses to food reward are unchanged in rats with neuropathic pain. Pain 2016;157:2731–2738. [PubMed: 27548047]

[42]. Parent AJ, Beaudet N, Beaudry H, Bergeron J, Bérubé P, Drolet G, Sarret P, Gendron L. Increased anxiety-like behaviors in rats experiencing chronic inflammatory pain. Behav Brain Res 2012;229:160–167. [PubMed: 22245257]

[43]. Pitzer C, La Porta C, Treede R-D, Tappe-Theodor A. Inflammatory and neuropathic pain conditions do not primarily evoke anxiety-like behaviours in C57BL/6 mice. Eur J Pain 2019;23:285–306. [PubMed: 30098102]

[44]. Racine M Chronic pain and suicide risk: A comprehensive review. Prog Neuropsychopharmacol Biol Psychiatry 2018;87:269–280. [PubMed: 28847525]

[45]. Raison CL, Capuron L, Miller AH. Cytokines sing the blues: inflammation and the pathogenesis of depression. Trends Immunol 2006;27:24–31. [PubMed: 16316783]

[46]. Ravenelle R, Santolucito HB, Byrnes EM, Byrnes JJ, Tiffany Donaldson S. Housing environment modulates physiological and behavioral responses to anxiogenic stimuli in trait anxiety male rats. Neuroscience 2014;270:76–87. [PubMed: 24713371]

[47]. Richards BL, Whittle SL, Buchbinder R. Antidepressants for pain management in rheumatoid arthritis. Cochrane Database Syst Rev 2011:CD008920.

[48]. Schwartz N, Temkin P, Jurado S, Lim BK, Heifets BD, Polepalli JS, Malenka RC. Chronic pain. Decreased motivation during chronic pain requires long-term depression in the nucleus accumbens. Science 2014;345:535–542. [PubMed: 25082697]

[49]. Sena ES, Currie GL, McCann SK, Macleod MR, Howells DW. Systematic Reviews and Meta-Analysis of Preclinical Studies: Why Perform Them and How to Appraise Them Critically. J Cereb Blood Flow Metab 2014;34:737–742. [PubMed: 24549183]

[50]. Sheahan TD, Siuda ER, Bruchas MR, Shepherd AJ, Mohapatra DP, Gereau RW, Golden JP. Inflammation and nerve injury minimally affect mouse voluntary behaviors proposed as indicators of pain. Neurobiol Pain 2017;2:1–12. [PubMed: 29075674]

[51]. Soliman N, Rice ASC, Vollert J. A practical guide to preclinical systematic review and meta-analysis. Pain 2020. doi:10.1097/j.pain.0000000000001974.

[52]. Tang NKY, Beckwith P, Ashworth P. Mental Defeat Is Associated With Suicide Intent in Patients With Chronic Pain. Clin J Pain 2016;32:411–419. [PubMed: 26201013]

[53]. Tappe-Theodor A, King T, Morgan MM. Pros and Cons of Clinically Relevant Methods to Assess Pain in Rodents. Neuroscience & Biobehavioral Reviews 2019;100:335–343. [PubMed: 30885811]

[54]. Urban R, Scherrer G, Goulding EH, Tecott LH, Basbaum AI. Behavioral indices of ongoing pain are largely unchanged in male mice with tissue or nerve injury-induced mechanical hypersensitivity. Pain 2011;152:990–1000. [PubMed: 21256675]

[55]. Vachon-Presseau E, Roy M, Martel M-O, Caron E, Marin M-F, Chen J, Albouy G, Plante I, Sullivan MJ, Lupien SJ, Rainville P. The stress model of chronic pain: evidence from basal cortisol and hippocampal structure and function in humans. Brain 2013;136:815–827. [PubMed: 23436504]

[56]. Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, Antonic A, Howells DW, Macleod MR. Meta-analysis of data from animal studies: A practical guide. Journal of Neuroscience Methods 2014;221:92–102. [PubMed: 24099992]

[57]. Viechtbauer W Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software 2010;36:1–48.

[58]. Violle N, Balandras F, Le Roux Y, Desor D, Schroeder H. Variations in illumination, closed wall transparency and/or extramaze space influence both baseline anxiety and response to diazepam in the rat elevated plus-maze. Behav Brain Res 2009;203:35–42. [PubMed: 19389429]

[59]. Walf AA, Frye CA. The use of the elevated plus maze as an assay of anxiety-related behavior in rodents. Nature Protocols 2007;2:322–328. [PubMed: 17406592]

[60]. Wall PM, Messier C. Methodological and conceptual issues in the use of the elevated plus-maze as a psychological measurement instrument of animal anxiety-like behavior. Neuroscience & Biobehavioral Reviews 2001;25:275–286. [PubMed: 11378181]

[61]. Wickham H ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Springer International Publishing, 2016 doi:10.1007/978-3-319-24277-4.

[62]. Wright SL. Limited Utility for Benzodiazepines in Chronic Pain Management: A Narrative Review. Adv Ther 2020;37:2604–2619. [PubMed: 32378069]

[63]. Zhou W, Jin Y, Meng Q, Zhu X, Bai T, Tian Y, Mao Y, Wang L, Xie W, Zhong H, Zhang N, Luo M-H, Tao W, Wang H, Li J, Li J, Qiu B-S, Zhou J-N, Li X, Xu H, Wang K, Zhang X, Liu Y, Richter-Levin G, Xu L, Zhang Z. A neural circuit for comorbid depressive symptoms in chronic pain. Nature Neuroscience 2019;22:1649–1658. [PubMed: 31451801]
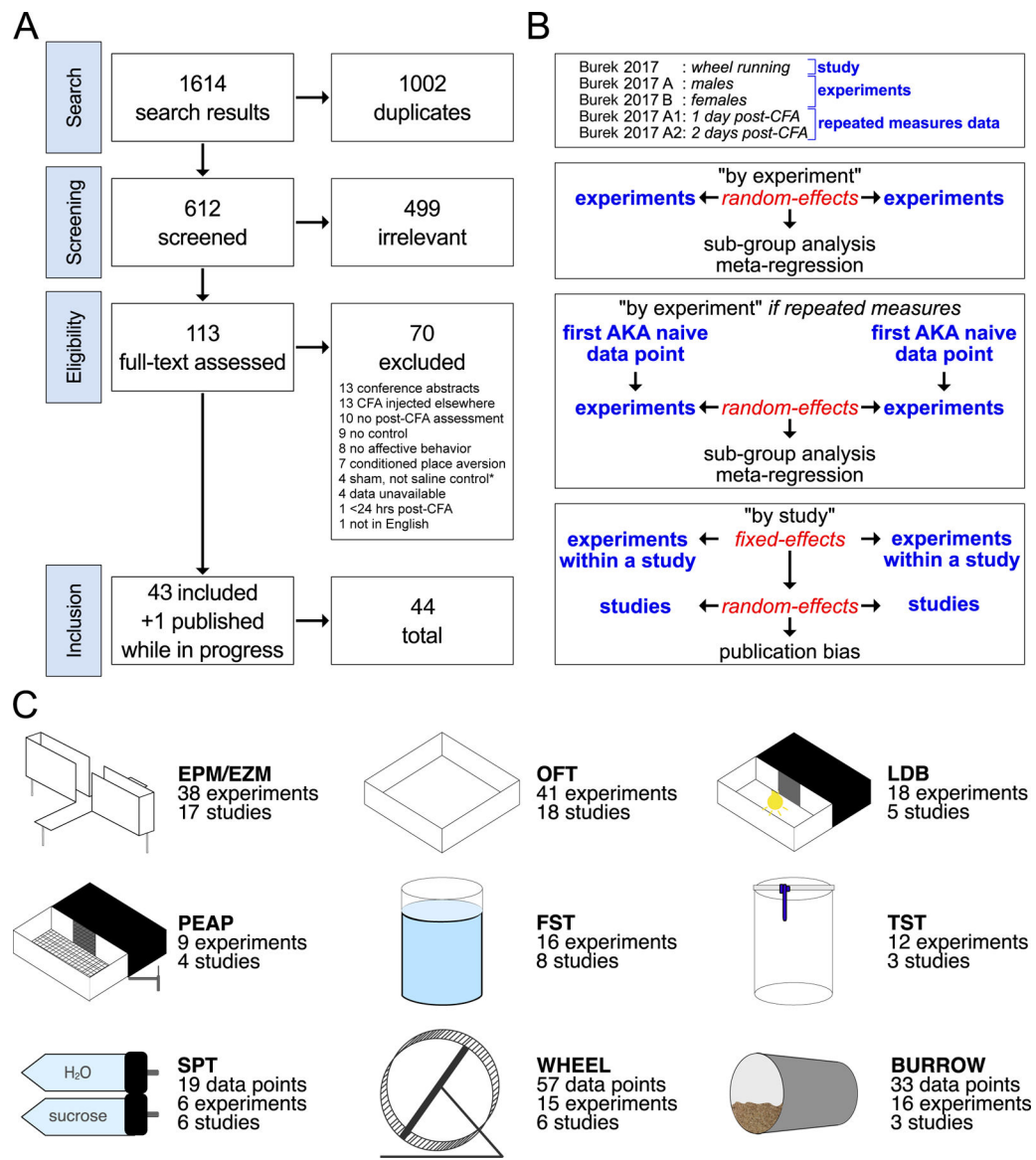
Figure 1.

Search strategy, analysis strategy, and results. **A**. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of search results, screening process, exclusion criteria, and final count of included studies. See methods for further information on "sham-not-saline" studies. **B**. Defined terminology and analysis strategy. **C**. Diagrams and metadata for the 9 behavioral tests included in this systematic review and meta-analysis: elevated plus/zero maze, open field test, light/dark box, place escape/avoidance paradigm, forced swim test, tail suspension test, sucrose preference test, wheel running, and burrowing. Main results include global summary effect estimates for all 9 tests and a subset of sub-group analysis and meta-regressions for the EPM/EZM, burrowing, and FST. Forest plots with individual experiments and studies, all sub-group analyses, and all meta-regressions per behavioral test are available in the supplemental file.
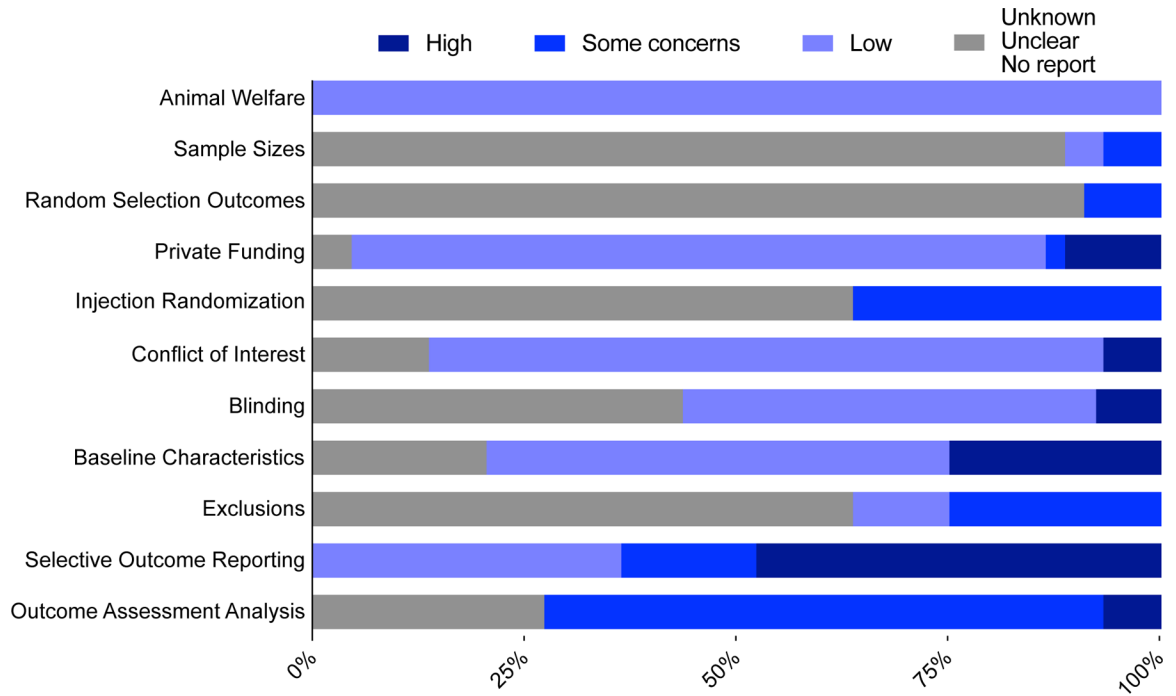
Figure 2.
Stacked bar chart of quality assessments on risks of bias. Darkest blue indicates high risk of bias, medium blue indicates some concerns, light blue indicates low risk of bias, and gray bars indicate percentage of studies for which no determination could be made. Rubric, adapted from the Systematic Review Centre for Laboratory Animal Experimentation's Risk of Bias tool, is detailed in the supplemental file.
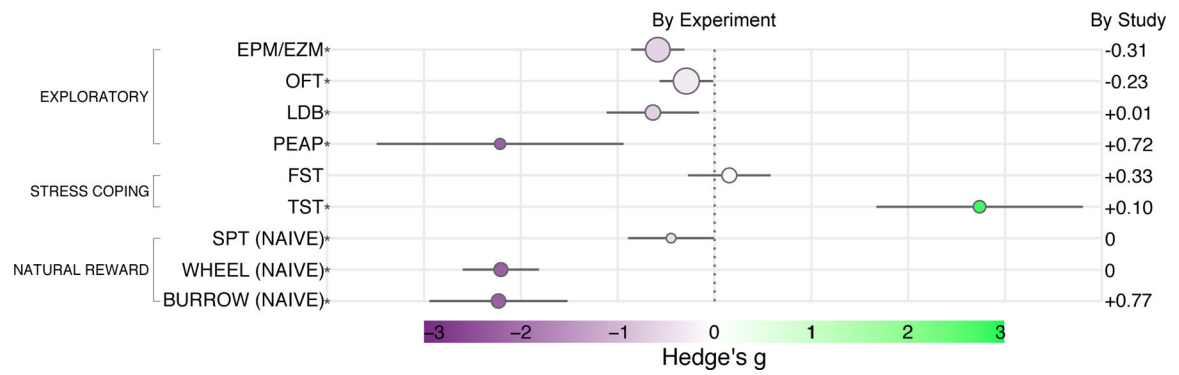
Figure 3.
Forest plot of global summary effect estimates as Hedge's *g* with 95% confidence intervals for each behavioral test. For the SPT, wheel running, and burrowing tests, the first data point of repeated measures, AKA when animals were naïve to repeated measures, is used. Behavioral test is abbreviated on the left with asterisks indicating significant effect. Size of circle indicates number of included experiments. Gray bars extending on either side of each circle delineate 95% confidence intervals. Dotted line marks the null effect. The color scale symmetrically extends from deep purple at *g* = −3.0 to bright green at *g* = 3.0, with white at *g* = 0. Difference between "by experiment" and "by study" global summary estimates is indicated on the right for reference. Individual forest plots by experiment and by study for all tests, and using naïve or data point with greatest effect for SPT, wheel running, and burrowing, are available in the supplemental file.
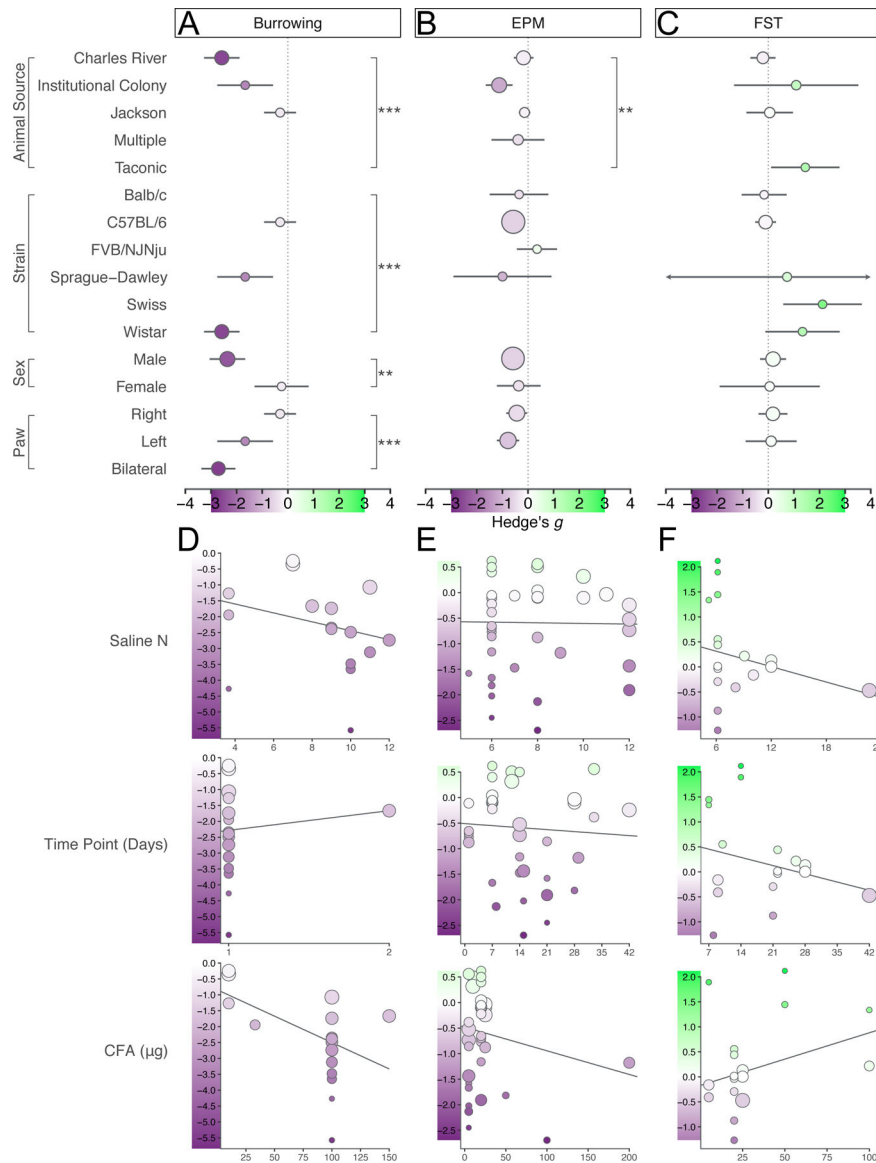
Figure 4.
Sub-group forest plots and meta-regression plots to explore potential sources of heterogeneity for burrowing, elevated plus/zero maze, and forced swim test. **A–C**. Effect sizes per subgroup plotted as circles with 95% confidence intervals marked by the underlying line. Size of circle represents the number of experiments belonging to that subgroup. Color of circle matches a symmetrical scale extending from the lowest effect size in darkest purple to the highest effect size in green, with zero as white. Color scale legend is overlaid on each x-axis and is consistent across the three panels. Dotted line provides a marker for x=0 across the y-axis. Empty spaces per panel indicate no experiments belonging to that sub-group. Sub-group membership overlaps. **D–F**. Effect size is on the y-axis and levels of each continuous variable moderator are on the x-axis. Circles represent effect size of individual experiments. Color and size scales are specific to the range of effect sizes for each behavioral test.
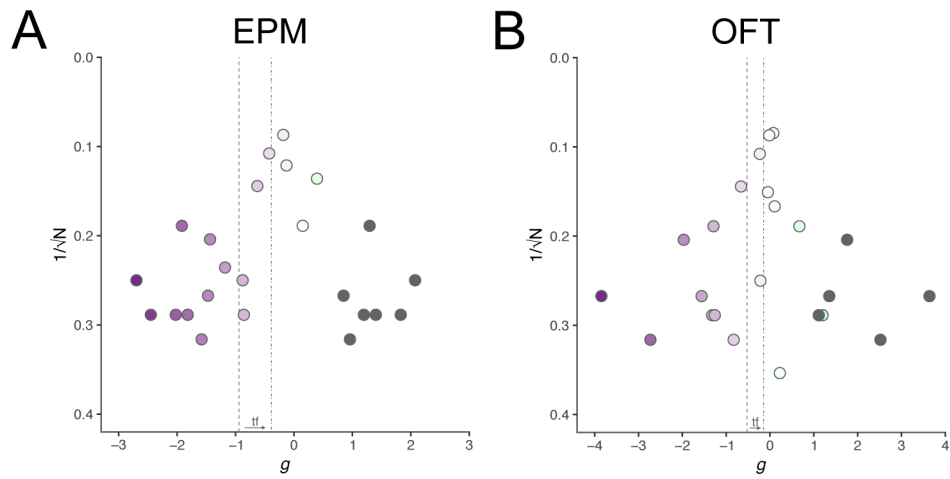
Figure 5.
Funnel plots for determining the extent of publication bias per behavioral test with trim-and-fill analysis. Circles indicate effect size for individual experiments. Color scales of circles are specific to each behavioral test. Gray circles indicate imputed experiments from trim-and-fill analysis. Dashed line indicates global summary effect estimate before trim-and-fill. Dashed-and-dotted line indicates the global summary effect estimate after trim-and-fill. Arrow and "tf" between the two lines near x-axis indicates direction of change.