Protocol for a systematic review and meta-analysis of data from preclinical studies employing forced swimming test: an update

A B Ramos-Hryb,1 Z Bahor,2 S McCann,2 E Sena,2 M R MacLeod,2 C Lino de Oliveira1

ABSTRACT

Objective Forced swimming test (FST) in rodents is a widely used behavioural test for screening antidepressants in preclinical research. Translational value of preclinical studies may be improved by appraisal of the quality of experimental design and risk of biases, which remains to be addressed for FST. The present protocol of a systematic review with meta-analysis aims to investigate the quality of preclinical studies employing FST to identify risks of bias in future publications. In addition, this protocol will help to determine the effect sizes (ES) for primary and secondary outcomes according to several aspects of the FST study design.

Search strategy, Screening annotation, Data management Publications reporting studies testing different classes of antidepressants in FST will be collected from Medline, SCOPUS and Web of Science databases. A broad list of inclusion criteria will be applied excluding those studies whereby FST is used as a stressor or studies reporting data from co-treatments. For assessing the quality of the included publications, the quality checklist adapted by Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies will be used. If the meta-analysis seems feasible, the ES and the 95% CI will be analysed. The heterogeneity between studies will be assessed by using the $\chi^2$ statistic with $n-1$ degrees of freedom. Subgroup meta-analysis (meta-regression, and if necessary, stratified regression) will be performed when possible according to characteristics of study design and study quality to assess their impact on efficacy of the treatments. In addition, funnel plotting, Egger regression, and ‘trim and fill’ will be used to assess the risk of publication bias. Results of this protocol will help to create rational methodological guidelines for application of FST in rodents and improve the quality and translational value of preclinical research on antidepressant discovery.

Reporting A preliminary version of the present protocol has been preregistered with Systematic Review Facility (http://syrf.org.uk/). A preprint version of the current protocol has been registered with Open Science Framework (https://osf.io/9kxm4/). Results will be communicated in scientific meetings and peer-reviewed journals. We plan to conduct an anonymous and online survey within the scientific community to ask researchers about their perception of risk of bias and their experience with the publication of negative results.

Strengths and limitations of this study

► This protocol for systematic review will collect, with broad inclusion criteria, preclinical studies employing forced swimming test (FST).
► The present protocol has been preregistered with Open Science Framework.
► A preliminary version of the present protocol has been preregistered with Systematic Review Facility (Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies).
► Results obtained with this systematic review and meta-analysis may help to create specific and rational methodological guidelines for application of FST in rodents.
► High levels of heterogeneity between studies may limit the external validity of our results.
► The summary effect size may be overestimated by publication bias.

INTRODUCTION

Major depression disorder (MDD) in humans is characterised by depressed mood and behavioural inhibition and often comes with social avoidance, generalised anxiety, eating disorders, sleeping and problems to cope with stress.1 Despite the difficulty in finding suitable models to mimic subjective, behavioural and neurobiological aspects of MDD, there are several animal models predictive of MDD treatment.2,3 Most of these animal models are based on behavioural responses of an animal to inescapable stress, providing a framework for several laboratory tests.4–7 Usually, inescapable stress induces behavioural inhibition (or immobility) that can be counteracted by antidepressant treatment.4–7 Therefore, behavioural tests in animals are employed as screening steps during the preclinical phase of antidepressant drug discovery. Forced swimming test (FST)4 in rats and mice is used in preclinical trials of antidepressants. FST is easy to run, inexpensive, sensitive and
relatively selective to known antidepressants (for review see Cryan et al\(^\text{b}\)). One criticism that may apply to FST is the abundance of 'positive results'\(^\text{9-11}\) that contrasts with the failure of antidepressant treatments in some clinical trials or therapeutics.\(^\text{9-11}\) There is an estimation that up to 50% of patients are resistant to the treatment with the antidepressants currently available.\(^\text{9-10}\) Many different reasons may account for the contrasting findings between preclinical and clinical data\(^\text{11}\) including individual variability, poor quality of the studies as well as publication bias. Publication bias in a preclinical field may inflate the estimated effect size (ES)\(^\text{12-13}\) leading to inflated expectations of efficacy in clinical trials, which may explain partially the perceived contrast between fields.\(^\text{31}\) Therefore, the aim of the present study is to evaluate the quality of published literature applying FST to detect effects of the treatment with antidepressants and the risk of bias in this research field.

Initially, a pilot study was performed to create a database and to standardise the methods for a systematic review.\(^\text{14-15}\) This pilot study started with a review in Medline and Embase retrieving more than 7000 publications by using expressions commonly found in the literature such as ‘forced swimming test’ OR ‘forced swim test’ OR ‘Porsolt test’ OR ‘fst’. The combination of these with medical subject heading (MeSH) terms related to ‘rodents’\(^\text{16}\) and ‘antidepressants’\(^\text{17}\) retrieved the publications more relevant for the present study. For screening purposes, a database containing bibliographical information from retrieved publications was built. We applied inclusion and exclusion criteria in the screening steps. Forty references, randomly selected from the database, generated 20 references to the pilot study, that is, one reference in every two fitted the inclusion criteria. From the selected literature, parameters were taken to estimate: (1) Quality. (2) ES. (3) Heterogeneity. (4) Publication bias. Most of the studies included in the pilot study were published from 2007 onwards. The quality score scale, adapted from the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies (CAMARADES),\(^\text{18-19}\) revealed that 12 studies scored above the median score (mean=09, maximum=14, minimum=04); none of them scored the maximum (18) or the minimum (0) values of the scale. Interestingly, all studies reported ‘species or strain in the title, abstract or full text’ and none reported ‘sample size calculation’ or ‘concealment of treatment allocation’ indicating that year of publication may not influence the quality of the studies. The median scores for studies in rats and mice were equal to the overall median score. Most of the experiments were performed in male animals. There were 34 different experiments: 15 using tricyclic antidepressants, 16 using selective serotonin reuptake inhibitors and 3 using selective norepinephrine reuptake inhibitors.

In the 34 experiments, there were 96 comparisons between experimental (439–470 animals) and control groups (276–287). In most of the mentioned experiments, a control group was compared with several different experimental groups generating significant and non-significant results in a single experiment. The incidences of significant results for primary and secondary outcomes within the experiments were 88.2% and 84.6%, respectively. Interestingly, the experiments also showing non-significant results for primary outcomes were only 29.4% whereas most of the experiments (92.3%) also reported non-significant results for secondary outcomes. The high number of significant results as compared with the negative ones was also found in another study.\(^\text{8}\) In summary, these preliminary analyses indicated that quality scores will be independent of the publication date, the experimental species as well as the type of antidepressant tested. The differences in sex, strains and ages may be a source of heterogeneity. In addition, it is expected that ‘random allocation to a treatment’, ‘concealment of treatment allocation’ and ‘sample size calculation’ will be neglected in this field of research. Moreover, these interim data suggest the existence of publication bias. Considering the sample used in the pilot study as the representative sample, the screening process may generate enough data (50% of publications in the database will fit inclusion criteria, ie, 2200 publications) for reliable estimation of the quality of the studies, ES, heterogeneity and publication bias in the field.

A preliminary version of this protocol was deposited in the Systematic Review Facility in February 2016.\(^\text{14-15}\) The current version of the protocol was updated based on procedures available in CAMARADES following instructions by de Vries et al\(^\text{20}\) A preprint version of the current text was preregistered with Open Science Framework (osf.io/9kxm4). Meta-analysis will be performed after publication of the protocol in a journal with a peer-review system. The links containing preliminary versions of the protocol will be updated to acknowledge the existence of the final peer-reviewed version.

**Systematic review questions**

► What is the quality level of these studies?
► Is there any relationship between quality scores and ES of outcomes reported in preclinical studies employing FST?
► Is there an influence of the study design or the ES in primary or secondary outcomes of these preclinical studies?
► Is there any risk of bias in preclinical studies employing FST for antidepressant research?

**METHODS AND ANALYSIS**

**Systematic review in specialised literature**

This protocol was formulated using the Systematic Review Center for Laboratory animal Experimentation format.\(^\text{20}\) The search strategy is based on previously reported protocols\(^\text{17}\) and consists of an update from our previous protocol registered in the CAMARADES platform.\(^\text{14-15}\) Medline, SCOPUS and Web of Science will be the databases selected. The search in Medline will be performed using the Pubmed platform (advanced search in http://www.ncbi.nlm.nih.gov).
The relevant period of publication started from 1977 when the first paper was published4 to December 2017.

Inclusion criteria
The following inclusion criteria will be applied to the systematic review outcome:

► Publication date: since 1977, the year the first paper was published, to present (December 2017).
► Language: any language.
► Animal species: rats and mice, regardless of age and sex.
Type of publication: all types of publications containing studies describing the effect of all classes of clinically tested antidepressant drugs in FST, compared with control animals treated or not with vehicle will be included, regardless of randomisation. The antidepressants included in this review will be dose listed in the protocol by McCann et al. in future studies, we intend to include publications containing information about candidate substances (plant-derived compounds such as polyphenols and terpenoids, and ketamine).

Studies with any route, dose and treatment schedule for drug administration are eligible.

**Exclusion criteria**

The following exclusion criteria will be applied to the inclusion criteria application outcome:

- Experiments using FST in rats or mice, only as a stressor, without showing the data of the behavioural measures.

**Experiments reporting data of co-treatments.** The publications containing these experiments will be kept if they also report experiments with single treatments.

**Search strategy**

Publications returned from the searches will be exported to a single reference manager file. Duplicate references will be deleted. Two investigators will independently evaluate the titles and abstracts obtained to assess if they meet the inclusion criteria and compare their results. If there is any discrepancy in included titles, consensus will be reached through discussion with a third investigator.

**Data extraction**

Data on outcome measures (primary and secondary) and attributes of study quality (see items 3.6, 3.7 and 3.8) will be recorded. One investigator will carry out initial data extraction and a second investigator will then check all data entered. Primary outcome measures extracted will be recorded. Primary outcome measures extracted will

---

**Table 2** Keywords for search in the SCOPUS database

| (rodentia OR mice) OR mus OR mouse) OR murine) OR woodhouse) OR rats) OR rat) OR murinae) OR muridae) OR cottonian) OR cottonrat) OR rodentia) OR rodent) OR rodents) #1 |
| ssri OR ‘selective serotonin-reuptake inhibitor’ OR fluoxetine OR citalopram OR escitalopram OR fluvoxamine OR paroxetine OR sertraline OR dapoxetine OR snri OR ‘serotonin and norepinephrine reuptake inhibitor’ OR duloxetine OR levomilnacipran OR sibutramine OR bicalifadine OR venlafaxine OR desvenlafaxine OR milnacipran OR tramadol OR TCA OR ‘tricyclic antidepressant’ OR amitriptyline OR butriptyline OR clomipramine OR desipramine OR dosulepine OR doxepin OR imipramine OR iprindole OR lofepramine OR miltiracen OR nortriptyline OR opipramol OR propriyptline OR trimipramine OR sari OR ‘serotonin antagonist and reuptake inhibitor’ OR etoperidone OR lorpiprazole OR mepiprazol OR lurasodone OR nefazodone OR trazadone OR nris OR ‘norepinephrine reuptake inhibitors’ OR atomoxetine OR reboxetine OR viloxazine OR ndri OR ‘norepinephrine dopamine reuptake inhibitor’ OR bupropion OR dexamethphenidate OR methylphenidate OR ndma OR nras OR ‘norepinephrine dopamine releasing agent’ OR amphetamine OR dextroamphetamine OR dextromethylamphetamine OR lisdexametamine OR teca OR tara) OR Texas OR ‘tetracyclic antidepressant’ OR amoxapine OR maprotiline OR mianserin OR mirtazaapine OR maoi OR ‘monoamine oxidase inhibitor’ OR isocarboxazid OR moclobemide OR phenelzine OR prilindol OR selegiline OR tranylcypromine OR antidepressant) #3 |

| (‘forced swimming test’ OR ‘forced swimming tests’ OR ‘forced swimming test’ OR ‘forced swim test’ OR ‘porsolt test’ OR ‘porsolt tests’) #3 |

---

**Table 3** Keywords for search in the Web of Science database

| (rodentia OR mice) OR mus) OR mouse) OR murine) OR woodhouse) OR rats) OR rat) OR murinae) OR muridae) OR cottonian) OR cottonrat) OR rodentia) OR rodent) OR rodents) #1 |
| ssri OR ‘selective serotonin-reuptake inhibitor’ OR fluoxetine OR citalopram OR escitalopram OR fluvoxamine OR paroxetine OR sertraline OR dapoxetine OR snri OR ‘serotonin and norepinephrine reuptake inhibitor’ OR duloxetine OR levomilnacipran OR sibutramine OR bicalifadine OR venlafaxine OR desvenlafaxine OR milnacipran OR tramadol OR TCA OR ‘tricyclic antidepressant’ OR amitriptyline OR butriptyline OR clomipramine OR desipramine OR dosulepine OR doxepin OR imipramine OR iprindole OR lofepramine OR miltiracen OR nortriptyline OR opipramol OR propriyptline OR trimipramine OR sari OR ‘serotonin antagonist and reuptake inhibitor’ OR etoperidone OR lorpiprazole OR mepiprazol OR lurasodone OR nefazodone OR trazadone OR nris OR ‘norepinephrine reuptake inhibitors’ OR atomoxetine OR reboxetine OR viloxazine OR ndri OR ‘norepinephrine dopamine reuptake inhibitor’ OR bupropion OR dexamethphenidate OR methylphenidate OR ndma OR nras OR ‘norepinephrine dopamine releasing agent’ OR amphetamine OR dextroamphetamine OR dextromethylamphetamine OR lisdexametamine OR teca OR tara) OR Texas OR ‘tetracyclic antidepressant’ OR amoxapine OR maprotiline OR mianserin OR mirtazaapine OR maoi OR ‘monoamine oxidase inhibitor’ OR isocarboxazid OR moclobemide OR phenelzine OR prilindol OR selegiline OR tranylcypromine OR antidepressant) #3 |

| (‘forced swimming test’ OR ‘forced swimming tests’ OR ‘forced swimming test’ OR ‘forced swim test’ OR ‘porsolt test’ OR ‘porsolt tests’) #3 |
be the parameters (total or mean duration, percentage or punctuation) of immobility. Secondary outcome measures will be the parameters of active behaviours (swimming, climbing and headshakes) and index of locomotion (in the open field, or rotarod or another test).

**Design of study**

Data on study design will be recorded, including: species, strain, age, weight and sex of animals used; number of experimental groups and number of animals per group; number of experiments and replications; housing conditions (food and water regimens, light cycle, temperature, size of the cage, length of housing the laboratory conditions); experimental conditions (time, illumination, dimensions of the tank, temperature and volume of water); FST protocol (single or repeated sessions, eg, only test, or pretest followed by one or more tests, length of swimming sessions); antidepressant subtype, dose (mg/kg) or regimen (single or multiple), mode of action of administered antidepressant; timing of drug administration related to the time point of the outcome measurement (test session of FST); methods of outcome measurement (manual or automatised); statistical methods for comparing groups and specific data from behavioural measures (mean, SD or SE of mean); reporting of data exclusion or inclusion for analysis.

**Quality of study**

A checklist with 10 items adapted from CAMARADES will be used to assess the quality of experimental design. Additionally, a user-defined checklist will be used to assess the quality of the protocols of FST. The items in the user-defined checklist were chosen considering previous experience of the authors (eg, Lino-de-Oliveira et al and Mezadri et al, and other published literature, eg, Porsolt et al, Detke et al and Petit-Demouliere et al). Information in the user-defined checklist is relevant to reproducibility of the study. Publications will receive a point for compliance of each item in the checklist from which group median scores will be calculated.

Ten items in the checklist adapted from CAMARADES are: (1) Peer-reviewed publication. (2) Reporting species/strain of animals in the title or abstract and in the full text. (3) Statement of compliance with animal welfare regulations. (4) Use of animals with behavioural phenotype. (5) Reporting blinded assessment of outcome. (6) Statement of possible conflicts of interest. (7) Reporting randomisation of treatment allocation. (8) Reporting sample size calculation. (9) Reporting criteria for inclusion and/or exclusion of data. (10) Reporting concealment of treatment allocation.

Ten items in the user-defined checklist are: (1) Reporting sex of animals in the title or abstract or in the full text. (2) Reporting age or weight or phase of development of animals in the title or abstract or in the full-text reporting. (3) Reporting efforts for improving animal welfare (eg, environmental enrichment) and reducing the number of animals. (4) Reporting control of temperature and light phase in the animal house. (5) Reporting dimensions of the tank used in FST. (6) Reporting the volume or the height of the water in the tank. (7) Reporting the temperature of water used in FST. (8) Reporting the use of clean water in FST. (9) Control for impaired locomotion of animals. (10) Reporting method of behavioural measurements.

**Outcome extraction**

The number of animals (N), the mean (M), SD or SE of the primary outcome (parameters of immobility) will be extracted for each treatment comparison. When available in the publications, N, M, SD or SE of secondary outcomes (ambulation in the open field test or active behaviours in FST) will be extracted. Ambulation in the open field test is used in many studies to discern influence of treatments on the motor function of animals (eg, while active behaviours in FST such as swimming and climbing may indicate mechanism of action of antidepressants in rats. In cases where the full data required for meta-analysis are not available in the text or tables in the publications, digital ruler software will be used to measure data from graphs. If information is completely unavailable in the publication, it will be requested directly from authors. When required data are not obtainable, such studies will be excluded from the analysis.

**Meta-analysis**

Analytical choices were made from pilot studies published elsewhere. Normalised mean difference will be used to calculate ES. SE of ES will be calculated for each comparison. In addition, the 95% CI of the ES will be calculated. Data will be aggregated using a weighted average method in which greater weight is given to more precise studies. For anticipated heterogeneity between studies, the random-effects model of DerSimonian and Laird will be used, which is more conservative than fixed-effect models, given the weighting towards more precise studies. For anticipated heterogeneity between studies, the random-effects model of DerSimonian and Laird will be used, which is more conservative than fixed-effect models, given the weighting towards more precise studies. For anticipated heterogeneity between studies, the random-effects model of DerSimonian and Laird will be used, which is more conservative than fixed-effect models, given the weighting towards more precise studies. For anticipated heterogeneity between studies, the random-effects model of DerSimonian and Laird will be used, which is more conservative than fixed-effect models, given the weighting towards more precise studies. For anticipated heterogeneity between studies, the random-effects model of DerSimonian and Laird will be used, which is more conservative than fixed-effect models, given the weighting towards more precise studies. For anticipated heterogeneity between studies, the random-effects model of DerSimonian and Laird will be used, which is more conservative than fixed-effect models, given the weighting towards more precise studies. For anticipated heterogeneity between studies, the random-effects model of DerSimonian and Laird will be used, which is more conservative than fixed-effect models, given the weighting towards more precise studies.
CONCLUSION

Problems in reproducibility of preclinical studies are one of the major concerns in antidepressant research. Importantly, medical decisions are grounded on clinical trials and these are based on results reported in preclinical studies. Therefore, the existence of publication or confirmation risk of bias as well as low statistical power and method quality prevent from obtaining accurate evidence for developing therapeutic interventions. Hence, results obtained with the present protocol for conducting a systematic review with meta-analysis of preclinical studies using FST for testing antidepressant responses in rodents intend to improve the quality and statistical power of future studies and to contribute towards applying the 3R (replacement, reduction and refinement) principles in preclinical research.

Acknowledgements The authors thank all members of the Camarades research group for their support in different aspects of this work.

Contributors CLdO selected the research theme, designed and performed research, analysed data, wrote the paper, approved the final version of the manuscript. ABR-H wrote the paper, approved the final version of the manuscript. ZB performed research, analysed data, discussed data, revised the paper, approved the final version of the manuscript. SM approved the final version of the manuscript. ES designed the research; discussed data, revised the paper, approved the final version of the manuscript.

Funding This research group received funds from CNPq, Capes, Fapesc (Brazilian public agencies for research), Alexander von Humboldt Foundation (Germany) and Newton International Exchanges (Royal Society, UK).

Competing interests None declared.

Open Practices

Data availability statement All data relevant to the study are included in the article.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

Open data Open data are available at https://osf.io/9KXM4/

Open peer review Prepublication and Review History is available online at http://dx.doi.org/10.1136/bmjoss-2017-000043.

REFERENCES


11. Belzung C. Innovative drugs to treat depression: did animal models fail to be predictive or did clinical trials fail to detect effects? Neuropsychopharmacology 2014;39:1041–51.


