

## PEER REVIEW HISTORY

BMJ Open Science publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://openscience.bmj.com/pages/wp-content/uploads/sites/62/2018/04/BMJ-Open-Science-Reviewer-Score-Sheet.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Systematic review and meta-analysis of data from preclinical studies employing the forced swimming test: an update
<b>AUTHORS</b>	Ramos-Hryb, A. B.; Bahor, Z., McCann, S. , Sena, E., MacLeod, M. R., Lino de Oliveira, C.

### VERSION 1 - REVIEW

<b>REVIEWER 1</b>	<i>H Einat</i> <i>The Academic College of Tel Aviv-Yaffo</i>
<b>REVIEW RETURNED</b>	09-06-18

<b>GENERAL COMMENTS</b>	<p>Recommendation Reject</p> <p>The current paper describes a protocol for a planned Systematic review and meta-analysis of data from preclinical studies employing the forced swimming test. Variations of this protocol were already published in Systemic Review Facility (Syrf, Camarades, <a href="http://syrf.org.uk/protocols/">http://syrf.org.uk/protocols/</a>) in 2016 and is publicly available in Open Science Framework. Moreover, some relevant information regarding the pilot exploration leading to this protocol and planned study was recently mentioned in a comment by these authors in Neuroscience and Biobehavioural Reviews (<a href="https://doi.org/10.1016/j.neubiorev.2018.05.025">https://doi.org/10.1016/j.neubiorev.2018.05.025</a>).</p> <p>Whereas the protocol described in the current MS is appropriate, I would suggest that it was already mentioned in a number of venues (although not precisely as described here) and I think that instead of describing the protocol, the authors may want to invest their time in data collection and in the study itself. In short, I will be looking forward for the results of the meta-analysis but I am not sure that another publication of a protocol is helpful in any way.</p> <p>Additional problems</p> <p>The authors suggest (in the abstract and introduction) that one reason to suspect publication bias is the discrepancy between success in animal models and failure in patients. Whereas in general this is a real problem, the analysis of the effects of standard antidepressants in the FST will not help us overcome such bias. This is because standard antidepressants are</p>
-------------------------	---

	<p>efficacious in patients in most studies (and all meta-analyses to the best of my knowledge). Accordingly, the fact that most studies of these antidepressants also show positive effects in the FST is not surprising. Indeed, many patients do not respond to standard antidepressants but this is not an issue of group statistics but of individual variability. A meta-analysis of group statistics will not help us to gain better understanding into issues of individual variability.</p> <p>Secondary measures: indeed swimming and climbing are additional measures in the FST (beyond immobility) but are clearly not independent and are directly related to the immobility time measures. Moreover, these measures were suggested to be related to the type of antidepressant used but that was validated more in rats than in mice. Additional measures suggested by the authors including locomotion or coordination measures are not clearly related to the behavior in the FST and therefore I am not sure what will they contribute to the meta-analysis.</p> <p>Online survey: The protocol suggests an online survey of researchers regarding their opinion on the issue of the risk of bias in the FST. With all due respect, I do not see the relevance of an online survey of opinions to real data. Such surveys are used in the clinical world at times to try to establish consensus statements when data is not enough and clinical decisions must be made but I do not see the relevance of such survey to a preclinical work where data is available. I would hope that scientific research would stay close to the data and away from opinions and beliefs.</p>
--	---

<b>REVIEWER 1</b>	<b>Rene Bernard</b> <b>Charité Universitätsmedizin Berlin</b>
<b>REVIEW RETURNED</b>	04-09-2018

<b>GENERAL COMMENTS</b>	<p>I had the pleasure to review the submitted study protocol from Lino de Oliveira et al. considered for publication in BMJ Open Science.</p> <p>The protocol is in line with the requirements for study protocols considered for publication in BMJ Open Science. However, one requirement is currently not fulfilled and needs to be updated: The current title does not reflect enough that this is a study protocol. Please include, as indicated by the requirements, the word protocol in the title.</p> <p>In the introduction, reference 8 is quoted in the following manner: <i>In the 34 experiments contrasted there were 96 comparisons between experimental groups (439-470 animals) and control groups (276-287). Significant results for primary and secondary outcomes were found in 88.2% and 84.6% of the</i></p>
-------------------------	---

	<p><i>experiments, respectively. Interestingly, non-significant results for primary outcome were observed only in 29.4% of the experiments whereas 92.3% reported non-significant results for secondary outcomes. The high number of significant results as compared to the negative ones were also found in another study.</i></p> <p>This leaves the reader slightly confused because significant and non-significant results do not sum up to 100%. How can secondary outcomes have 84.6% significant and 92.3% non-significant results at the same time. Please explain and change the wording or only quote the part you want to emphasize without confusing the reader.</p> <p>The protocol itself is well written and clearly explained and the outcome is of general interest and informative. However, one limitation is not noted or addressed: The authors try to link the poor translation results with positive outcome of the FST experiments. As quality criteria, the authors use 18 self-selected parameters and judge the quality of the study and therefore the results by the presence or absence of these criteria. Since journals or research societies have no recommendations or reporting guidelines for FST experiments, the absence or presence of some particular items may not be sufficient to make conclusions about the quality study, such as reporting on control of temperature and light phase (5), reporting temperature and cleaning of the water (4), reporting actions for improving animal welfare (16), statement of possible conflicts of interest (18). These items might characterize the excellence in reporting but do not address whether the results are reproducible.</p> <p>Therefore, I strongly suggest a modification in the analysis of the results: grouping of parameters that reflect descriptive reporting of the study on the one side; and contrast them with items that clearly address bias on the conduct, analysis and planning of study and therefore are more directly related to reproducibility; e.g. items 1, 9, 10, 11, 12, 14, 17.</p> <p>If this distinction is not made an unweighted analysis of all quality criteria together might limit the interpretation of the study.</p> <p>Sincerely, René Bernard</p>
--	---

**VERSION 1 – AUTHOR RESPONSE**

Manuscript bmjos-SP-1514661972

Response to reviewers:

REVIEWER 1:

Dear Dr. Einat: thank you for spending your time criticizing and improving our manuscript. We agreed with most of the points you made in the review. We split your remarks into five different categories and tried to explain our views on them. Please see our responses below and respective modifications in the reviewed manuscript. We are available for any other information you may require. Sincerely, Authors.

COMMENT#1: The current paper describes a protocol for a planned Systematic review and meta-analysis of data from preclinical studies employing the forced swimming test. Variations of this protocol were already <https://mail.google.com/mail/u/0?ik=102ffbea62&view=pt&search=all&permthid=thread-f%3A1611528719932398840&simpl=msg-f%3A16115287...> published in Systemic Review Facility (Syrf, Camarades, <http://syrf.org.uk/protocols/>) in 2016 and is publicly available in Open Science Framework.

RESPONSE #1: Thank you for your remark on this subject. As you mentioned, this protocol is already available in two public platforms (Syrf and OSF). We added this information in the topic “**Ethics and dissemination**” of the abstract. Syrf has a public repository for protocols of systematic reviews and meta-analysis for animal studies in the Syrcle format. OSF is a broad platform to deposit projects, material related to projects and preprints. Although Syrf and OSF represent transparent ways for sharing information, we consider also important to receive feedback from our peers before to start the exhausting work of extracting data for meta-analysis. Therefore, we decided submitting the protocol for a journal with a peer-review system. Fulfilling our expectations, we received important feedback to improve our work. We added an explanatory text in the end of the Introduction (please the see reviewed manuscript).

COMMENT#2: Moreover, some relevant information regarding the pilot exploration leading to this protocol and planned study was recently mentioned in a comment by these authors in Neuroscience and Biobehavioural Reviews (<https://doi.org/10.1016/j.neubiorev.2018.05.025>). Whereas the protocol described in the current MS is appropriate, I would suggest that it was already mentioned in a number of venues (although not precisely as described here) and I think that instead of describing the protocol, the authors may want to invest their time in data collection and in the study itself. In short, I will be looking forward for the results of the meta-analysis but I am not sure that another publication of a protocol is helpful in any way.

RESPONSE #2: Thank you for making this point. As you mentioned, we recently referenced the protocols available in Syrf (preliminary version, <http://syrf.org.uk/>) and OSF (preprint format of the current version, [osf.io/9kxm4](https://osf.io/9kxm4)) in a comment published in *Neuroscience and Biobehavioural Reviews* (Ramos-Hryb et al., 2018, <https://doi.org/10.1016/j.neubiorev.2018.05.025>). It worth to mention that interim data presented in the commentary (Ramos-Hryb et al., 2018) were from a small, exploratory, pilot study including only eleven publications, seventeen experiments and 259 mice treated with imipramine or vehicle. After publication in a journal with a peer-review system, the contents in the links mentioned above will be updated to inform the existence (hopefully) of the peer-reviewed version. We added an explanatory text in the end of the Introduction (please the see reviewed manuscript).

COMMENT #3: The authors suggest (in the abstract and introduction) that one reason to suspect publication bias is the discrepancy between success in animal models and failure in patients. Whereas in general this is a real problem, the analysis of the effects of standard antidepressants in the FST will not help us overcome such bias. This is because standard antidepressants are efficacious in patients in most studies (and all meta- analyses to the best of my knowledge). Accordingly, the fact that most studies of these antidepressants also show positive effects in the FST is not surprising. Indeed, many patients do not respond to standard antidepressants but this is not an issue of group statistics but of individual variability. A meta-analysis of group statistics will not help us to gain better understanding into issues of individual variability.

RESPONSE #3: Thank you for sharing your opinion, we agree there are uncountable levels between results of animal and clinical studies. In addition, we understand that meta-analysis is not suitable to address individual differences in animals or patients. Therefore, we decided to suppress it from the abstract where there is no place for further explanations. However, we kept the affirmation in the Introduction because publication bias in preclinical field may inflate the estimated effect size in animal models (refs. 12, 13) increasing the expectations of efficacy in clinical trials, which could explain the findings perceived as contrasting (e.g. ref 11). To make the point clearer, we added the following text to Introduction: “Many different reasons may account for the contrasting findings between preclinical and clinical data<sup>11</sup> 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<sup>12 13</sup> leading to inflated expectations of efficacy in clinical trial, which may explain partially the perceived

contrast between fields<sup>11</sup>.” Accordingly, we added references 12 and 13 in the reference list (please see the reviewed manuscript).

COMMENT #4: Secondary measures: indeed, swimming and climbing are additional measures in the FST (beyond immobility) but are clearly not independent and are directly related to the immobility time measures. Moreover, these measures were suggested to be related to the type of antidepressant used but that was validated more in rats than in mice. Additional measures suggested by the authors including locomotion or coordination measures are not clearly related to the behaviour in the FST and therefore I am not sure what will they contribute to the meta-analysis.

RESPONSE #4: Your analysis on this subject is solid and straightforward. We understand that our explanation was not sufficiently clear to justify our choice for secondary outcomes. Therefore, we modified the text of the outcome extraction (please see the reviewed manuscript).

COMMENT #5: Online survey: The protocol suggests an online survey of researchers regarding their opinion on the issue of the risk of bias in the FST. With all due respect, I do not see the relevance of an online survey of opinions to real data. Such surveys are used in the clinical world at times to try to establish consensus statements when data is not enough, and clinical decisions must be made but I do not see the relevance of such survey to a preclinical work where data is available. I would hope that scientific research would stay close to the data and away from opinions and beliefs.

RESPONSE #5: Your analysis on this subject is solid and straightforward. We understand that our explanation was not sufficiently clear to justify our survey. We also agree that an online survey could lead to inform opinions and beliefs however, we consider important that scientists in the field should share their experiences concerning negative results using the FST. With feedback from scientists using the FST, we hope to estimate the number of experiments with negative results that remain unpublished. Although we have published some of our negative data (<https://doi.org/10.1017/neu.2017.33>), we still have negative results unpublished. Therefore, we modified the text in the abstract (please see “Ethics and dissemination” in reviewed manuscript).

REVIEWER 2:

Dear Dr Bernard: thank you for spending your time criticizing and improve our manuscript. We agreed with most of the points you made in the review. We split your remarks into four different categories and tried to explain our views on them. Please see our responses below and respective modifications in the reviewed manuscript. We are available for any other information you may require. Sincerely, Authors.

COMMENT #1: I had the pleasure to review the submitted study protocol from Lino de Oliveira et al. considered for publication in BMJ Open Science. The protocol is in line with the requirements for study protocols considered for publication in BMJ Open Science. However, one requirement is currently not fulfilled and needs to be updated: The current title does not reflect enough that this is a study protocol. Please include, as indicated by the requirements, the word protocol in the title.

RESPONSE #1: Thank you for your kind remarks, we added the word protocol in the title (please see in reviewed manuscript).

COMMENT #2: In the introduction, reference 8 is quoted in the following manner: In the 34 experiments contrasted there were 96 comparisons between experimental groups (439-470 animals) and control groups (276-287). Significant results for primary and secondary outcomes were found in 88.2% and 84.6% of the experiments, respectively. Interestingly, non-significant results for primary outcome were observed only in 29.4% of the experiments whereas 92.3% reported non-significant results for secondary outcomes. The high number of significant results as compared to the negative ones were also found in another study. This leaves the reader slightly confused because significant and non-significant results do not sum up to 100%. How can secondary outcomes have 84.6% significant and 92.3% non-significant results at the same time. Please explain and change the wording or only quote the part you want to emphasize without confusing the reader.

RESPONSE #2: We agree that these paragraphs need clarification. Please see the following text and the revised manuscript: "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 to 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.”

COMMENT #3: The protocol itself is well written and clearly explained and the outcome is of general interest and informative. However, one limitation is not noted or addressed: The authors try to link the poor translation results with positive outcome of the FST experiments.

RESPONSE #3: We agree with your evaluation. Therefore, we deleted the statement on translational values from the list of strengths of the manuscript. (Please see in reviewed manuscript).

COMMENT #4: As quality criteria, the authors use 18 self-selected parameters and judge the quality of the study and therefore the results by the presence or absence of these criteria. Since journals or research societies have no recommendations or reporting guidelines for FST experiments, the absence or presence of some particular items may not be sufficient to make conclusions about the quality study, such as reporting on control of temperature and light phase (5), reporting temperature and cleaning of the water (4), reporting actions for improving animal welfare (16), statement of possible conflicts of interest (18).

RESPONSE #4: We agree with you and we know that there are no pre-existing guidelines that could be applied to evaluate these preclinical studies. However, based on our experience and considering several reviews that addressed this issue (e.g. ref 21 Petit-Demouliere *et al.* 2005), we considered each of the items in the checklist important to address the quality of the studies. Following your suggestion, we now divided the items in two different checklists (CAMARADES and User defined) addressing “different qualities” of the studies (please see in reviewed manuscript).

COMMENT #5: These items might characterize the excellence in reporting but do not address whether the results are reproducible. Therefore, I strongly suggest a modification in the analysis of the results: grouping of parameters, that reflect descriptive reporting of the study on the one side; and contrast them with items that clearly address bias on the conduct, analysis and planning of study and therefore are more directly related to reproducibility; e.g. items 1, 9, 10, 11, 12, 14, 17. If this distinction is not made an unweighted analysis of all quality criteria together might limit the interpretation of the study.

RESPONSE #5: We thank the reviewer for this very interesting suggestion and included it in our analysis (please see in reviewed manuscript).





# BMJ Open Science

BMJ Open Science is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open Science is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://openscience.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmj@bmj.com](mailto:info.bmj@bmj.com)

**Title:** Systematic review and meta-analysis of data from preclinical studies employing the forced swimming test: an update.

**Authors:** Ramos-Hryb, A. B.<sup>1</sup>; Bahor, Z.<sup>2</sup>, McCann, S.<sup>2</sup>, Sena, E.<sup>2</sup>, MacLeod, M. R.<sup>2</sup>, Lino de Oliveira, C.<sup>1</sup>

<sup>1</sup>Laboratory of Behavioral Neurobiology. CFS-CCB-UFSC- Florianópolis-SC-Brazil.

<sup>2</sup>Camarades-University of Edinburgh, Scotland, United Kingdom.

**Contact person and correspondence address:**

Dr. Cilene Lino de Oliveira

Departamento de Ciências Fisiológicas, Centro de Ciências Biológicas, Universidade Federal de Santa Catarina, Campus Universitário Trindade, 88049-900 – Florianópolis – SC – Brazil. Phone: +55 48 3721 4689, Fax: +55 48 3721 9672, E-mail: cilene@ccb.ufsc.br OR labnec.contato@gmail.com

## **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- This protocol for systematic review will collect, with broad inclusion criteria, preclinical studies employing the FST.
- The present protocol has been preregistered with Open Science Framework and a preliminary version was preregistered with Systematic Review Facility (Syrf, Camarades).
- Results obtained with this systematic review and meta-analysis may help to create specific and rational methodological guidelines for application of FST in rodents improving the quality and translational value of preclinical research on antidepressant discovery.
- High levels of heterogeneity between studies may limit the external validity of our results.
- The summary effect size may be overestimated by Publication bias.

## ABSTRACT

**Introduction:** Forced swimming test (FST) in rodents is a widely used behavioral test for screening antidepressants in preclinical research. However, there is a contrast between the high levels of “positive outcomes” reported in preclinical studies and the low efficacy of antidepressants in the clinics. Poor translational value of preclinical studies may be related to the presence of publication and confirmation risk of bias as well as with low quality of experimental design in the studies employing the FST, which remains to be addressed.

**Objective:** The present protocol of a systematic review with meta-analysis aims to investigate the quality of preclinical studies employing the FST to identify risks of bias in the future publications. In addition, this protocol will help to determine the effect sizes for primary and secondary outcomes according to several aspects of the FST study design.

**Methods and analysis:** Publications reporting studies testing different classes of antidepressants in the 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 the FST is used as a stressor or studies reporting data from co-treatments. For assessing the quality of the included publications, it will be used the CAMARADES’s adapted quality checklist. If the meta-analysis seems feasible, the effect size and the 95% confidence interval will be analyzed. The heterogeneity between studies will be assessed by using the Chi-square statistic with  $n-1$  degrees of freedom. Subgroup meta-analysis (meta-regression, and if necessary, stratified) 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, the funnel plotting, Egger regression and “trim and fill” will be used to assess the risk of publication bias. Moreover, we will conduct an anonymous and online survey within the scientific community where it will be asked to researchers about their perception of risk of bias, their use and reporting of measures against risk of bias, publication and communication of negative results. 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.

**Ethics and dissemination:** Results will be communicated in conferences, congresses, seminars, and peer-reviewed journal.

**KEYWORDS:** antidepressants, animal models, preclinical, bias risk.



## **1. INTRODUCTION**

Major depression disorder (MDD) in humans is characterized by depressed mood and behavioral inhibition and often come with social avoidance, generalized anxiety, disorders of eating, sleeping and problems to cope with stress<sup>1</sup>. Despite the difficulty in finding suitable models to mimic subjective, behavioral and neurobiological aspects of MDD, there are several animal models predictive of MDD treatment<sup>2 3</sup>. Most of these animal models are based on behavioral responses of an animal to inescapable stress, providing a framework for several laboratory tests<sup>4-7</sup>. Usually, inescapable stress induces behavioral inhibition (or immobility) that can be counteracted by antidepressant treatment<sup>4-7</sup>. Therefore, behavioral tests in animals are employed as screening steps during the preclinical phase of an antidepressant drug discovery. The Forced swimming test (FST<sup>4</sup>) 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<sup>2</sup>). However, one criticism that may apply to FST is the abundance of “positive results”<sup>8</sup> that contrasts with the failure of antidepressant treatments in clinical trials or in therapeutics<sup>9 10</sup>. There is an estimation that up to 50% of patients are resistant to the treatment with the antidepressants currently available<sup>9 10</sup>. Many different reasons may account for the contrasting findings between preclinical and clinical data<sup>11</sup> including factors disrupting the quality of preclinical or clinical studies. The aim in the present study is to evaluate the quality of the published literature reporting the use of FST to detect effects of the treatment with antidepressants.

Initially, a pilot study was performed to create a database and to standardize the methods for a systematic review (Pilot Study 1, See Supplementary material). This pilot study started with a review in the 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 terms with Mesh terms related to “rodents”<sup>12</sup> and “antidepressants”<sup>13</sup> retrieved the publications more relevant for the present study. For screening purposes, a database containing bibliographical information from retrieved publications was built and inclusion and exclusion criteria applied to select relevant studies. Forty references, randomly selected from the database, generated 20 references to the pilot study, i.e., 01 reference in every 02 fitted the inclusion criteria. From the selected literature, parameters were taken to estimate: 1- quality; 2- effect size, 2-heterogeneity, and 3-publication bias.

Most of the studies included in the pilot study were published from 2007 onwards. The quality score scale, adapted from CAMARADES<sup>14 15</sup>, revealed that twelve studies scored above the median score (median=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 study 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 rat and mice were equal to overall median score. Most of the experiments were performed in male animals. There were 34 different experiments: 1-fifteen using tricyclic antidepressants, 2- sixteen using selective inhibitors of serotonin reuptake (SSRI), 3- three using selective inhibitors of noradrenaline reuptake (SNRI). In the 34 experiments contrasted there were 96 comparisons between experimental groups (439-470 animals) and control groups (276-287). Significant results for primary and secondary outcomes were found in 88.2% and 84.6% of the experiments, respectively. Interestingly, non-significant results for primary outcome were observed only in 29.4% of the experiments whereas 92.3% reported non-significant results for secondary outcomes. The high number of significant results as compared to the negative ones were also found in another study<sup>8</sup>.

In summary, this preliminary analysis indicated that quality scores will be independent of the publication date, of the species of the experimental as well as of the type of antidepressant tested. The differences in sex, strains, and ages may be a source of heterogeneity. In addition, it is expected that experiments in females, “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 pilot study as representative, the screening process may generate enough data (50% of publications in the database will fit inclusion criteria, i.e., 2200 publications) to the reliable estimation of the quality of the studies, effect size, heterogeneity, and publication bias in the field. Hence, this protocol aims to assess the impact of methodological quality and publication bias on reported outcomes of FST. A preliminary version of this protocol was previously published in the Systematic Review Facility (Syrf) in February 2016<sup>16 17</sup> and updated based on procedures available in CAMARADES (Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies)<sup>13</sup> following instructions by de Vries et al.<sup>18</sup>. The current version was preregistered with Open Science Framework (DOI 10.17605/OSF.IO/9KXM4).

### **Systematic review questions:**

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

## **2. METHODS AND ANALYSIS**

### **2.1 Systematic review in specialized literature:**

This protocol was formulated using the SYRCLE (Systematic Review Center for Laboratory animal Experimentation) format<sup>18</sup>. The search strategy is based on previously reported protocols<sup>19</sup> and consists of an updating from our previously protocol registered in CAMARADES' platform<sup>16 17</sup>. Medline, SCOPUS, and Web of Science will be the database selected. The search in Medline will be performed using the platform Pubmed (Advanced search in <http://www.ncbi.nlm.nih.gov/pubmed>). The search in SCOPUS and Web of Science will be conducted accessing the platform "Periodicos CAPES" (Advanced search in <http://www-periodicos-capes-gov-br.ez46.periodicos.capes.gov.br/>) at the Federal University of Santa Catarina. The selection of Keywords was based on the different denominations of Forced Swimming Test (FST) found in the literature (see Supplementary Material). We decided to include in the review only the data from studies in rats and mice that are the most common laboratory species submitted to FST using the MeSH terms by Hooijmans et al.<sup>12</sup>. The list of antidepressants included in the research was by McCann et al.<sup>13</sup>. The relevant period of Publication date started in 1977 when the first paper was published<sup>4</sup> up to date (December 2017).

### **2.2 Inclusion criteria:**

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

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 to control animals treated or not with vehicle will be included, regardless of randomization. The antidepressants included in this review will be those listed in the protocol by McCann et al.<sup>13</sup>. 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.

### **2.3 Exclusion criteria:**

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

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

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

### **2.4 Search strategy**

The publication 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 broad inclusion criteria and compare their results. If there is any discrepancy in included titles, consensus will be reached through discussion with a third investigator.

### **2.5 Data extraction**

Data on outcome measures (primary and secondary) and attributes of study quality (see items 2.6, 2.7 and 2.8) will be recorded. One investigator will carry out initial data extraction and a second investigator will then check all data entered. Primary outcome extracted will be the parameters (total or mean duration, percentage or punctuation) of immobility. Secondary outcome measures will be the parameters of active behaviors (swimming, climbing, and headshakes) and index of locomotion (in the open field, or rota rod, or another test).

## **2.6 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, e.g. only test, or pre-test 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 automatized); statistics methods for comparing groups and specific data from behavioural measures (mean, standard deviation or standard error of mean); reporting of data exclusion or inclusion for analysis.

## **2.7 Quality of study**

Study quality will be assessed according to a quality checklist with 18 points in total (receiving 1 point for compliance of each item), adapted from CAMARADES<sup>14 15</sup>, for which the group median scores will be calculated: (1) peer reviewed publication; (2) reporting species/strain and age of animals in the title or abstract and in the full text; (3) reporting of dimensions of the tank and volume water; (4) reporting temperature and cleaning of water; (5) reporting control of temperature and light phase; (6) control for impaired locomotion of animals; (7) reporting method of behavioural measurements; (8) use of animals with co-morbidities; (9) reporting of sample size calculation; (10) randomization of treatment allocation; (11) concealment of treatment allocation; (12) blinded assessment of outcome and/or treatment or allocation of animals to groups; (13) statement of compliance with animal welfare regulations; (14) reporting sample size, mean and standard deviation or standard error of mean; (15) appropriate use of phenotype and animal model; (16) reporting actions for improving animal welfare (e. g. environmental enrichment); (17) reporting inclusion and/or exclusion data criteria; and (18) statement of possible conflicts of interest.

## **2.8 Outcome extraction**

The number of animals, the standard deviation (SD) or standard error (SE) of a mean outcome (parameters of immobility) will be extracted for each treatment comparison. Measures of active behaviors (swimming and climbing) will be extracted. In cases where the full data required for meta-analysis are not available from abstracts or publications, they will be requested directly from authors, and if unavailable, digital ruler software will be used to measure data from graphs. When required data are not obtainable, such studies will be excluded from the analysis.

## **2.9 Statistical analysis**

Normalized mean difference will be used to calculate the effect size (ES). SE of ES will be calculated for each comparison. In addition, the 95% confidence interval 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 model, given the weighting towards individual comparisons depends on the variance within those comparisons and on overall heterogeneity. The heterogeneity between studies will be assessed by using the Chi-square statistic with  $n-1$  degrees of freedom (df). To allow for multiple comparisons, a significance level will be set using Bonferroni correction taking into the number of comparisons. Publication bias will be looked for using funnel plotting, Egger regression and “trim and fill”. Subgroup meta-analysis (meta-regression, and if necessary, stratified) will be performed when possible according to characteristics of study design and study quality to assess their impact on efficacy, as the following subgroups: species, strain, age, co-morbidities (present/absent), FST protocol; antidepressant subtype, dose (mg/kg), regimen (single or multiple) and time of administration (relative to the test), time of outcome assessment (relative to the test), scores of quality checklist, especially the use of randomization and blinded assessment of outcome<sup>13 16 17</sup>.

## **3. 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 on preclinical studies. Therefore, the existence of publication or confirmation risk of bias as well as low statistical power and method quality prevents from obtaining accurate evidence for developing therapeutic

interventions. Hence, results obtained with the present protocol for conducting systematic review with meta-analysis of preclinical studies using the FST for testing antidepressant responses in rodents intends not only to improve the quality and statistical power of future studies but also to contribute in applying the principles of 3Rs in preclinical research.

### **Author Contributions**

CLO selected the research theme, designed and performed research, analyze data, wrote the paper, approved the final version of the manuscript.

ABRH wrote the paper, approved the final version of the manuscript.

ZB performed research, analyze data, discussed data, revised the paper, approved the final version of the manuscript.

SM revised the paper, approved the final version of the manuscript.

ES designed the research; discussed data, revised the paper, approved the final version of the manuscript.

MRM designed the research; discussed data, revised the paper, approved the final version of the manuscript.

### **Acknowledgments**

Authors thank all members of Camarades research group for their support in different aspects of this work.

### **Data availability statement**

The protocol is registered with Open Science Framework and all data retrieved and collected with this project will be available in [osf.io/9kxm4](https://osf.io/9kxm4) ((DOI 10.17605/OSF.IO/9KXM4) including Supplementary Material. A preliminary version of the present protocol has been preregistered with CAMARADES Protocols for Systematic Reviews of Animal Studies (RRID:SCR\_008970).

### **Funding statement**

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 statement**

The authors declare no known conflict of interests.

## **5. REFERENCES**

1. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Med* 2015;13:72. doi: 10.1186/s12916-015-0325-4 [published Online First: 2015/04/06]
2. Cryan JF, Valentino RJ, Lucki I. Assessing substrates underlying the behavioral effects of antidepressants using the modified rat forced swimming test. *Neurosci Biobehav Rev* 2005;29(4-5):547-69. doi: 10.1016/j.neubiorev.2005.03.008
3. Berton O, Durand M, Aguerre S, et al. Behavioral, neuroendocrine and serotonergic consequences of single social defeat and repeated fluoxetine pretreatment in the Lewis rat strain. *Neuroscience* 1999;92(1):327-41. [published Online First: 1999/07/07]
4. Porsolt RD, Bertin A, Jalfre M. Behavioral despair in mice: a primary screening test for antidepressants. *Arch Int Pharmacodyn Ther* 1977;229(2):327-36. [published Online First: 1977/10/01]
5. Detke MJ, Rickels M, Lucki I. Active behaviors in the rat forced swimming test differentially produced by serotonergic and noradrenergic antidepressants. *Psychopharmacology (Berl)* 1995;121(1):66-72.
6. Lino-de-Oliveira C, De Lima TC, de Pádua Carobrez A. Structure of the rat behaviour in the forced swimming test. *Behav Brain Res* 2005;158(2):243-50. doi: 10.1016/j.bbr.2004.09.004
7. Mezadri TJ, Batista GM, Portes AC, et al. Repeated rat-forced swim test: reducing the number of animals to evaluate gradual effects of antidepressants. *J Neurosci Methods* 2011;195(2):200-5. doi: 10.1016/j.jneumeth.2010.12.015 [published Online First: 2010/12/16]
8. 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. *Neurosci Biobehav Rev* 2018;84:1-11. doi: 10.1016/j.neubiorev.2017.11.003 [published Online First: 2017/11/09]
9. Carvalho AF, Cavalcante JL, Castelo MS, et al. Augmentation strategies for treatment-resistant depression: a literature review. *J Clin Pharm Ther* 2007;32(5):415-28. doi: 10.1111/j.1365-2710.2007.00846.x
10. Santaguida P, United States. Agency for Healthcare Research and Quality., McMaster University. Evidence-based Practice Center. Treatment for depression after unsatisfactory response to SSRIs. Comparative effectiveness review no 62. Rockville, MD: Agency for Healthcare Research and Quality, 2012:1 online resource (PDF file).
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(5):1041-51. doi: 10.1038/npp.2013.342 [published Online First: 2013/12/18]
12. Hooijmans CR, Tillema A, Leenaars M, et al. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. *Lab Anim* 2010;44(3):170-5. doi: 10.1258/la.2010.009117 [published Online First: 2010/06/15]
13. McCann SK, Irvine C, Mead GE, et al. Efficacy of antidepressants in animal models of ischemic stroke: a systematic review and meta-analysis. *Stroke* 2014;45(10):3055-63. doi: 10.1161/STROKEAHA.114.006304 [published Online First: 2014/09/02]

14. Macleod MR, O'Collins T, Howells DW, et al. Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke* 2004;35(5):1203-8. doi: 10.1161/01.STR.0000125719.25853.20 [published Online First: 2004/04/01]
15. Jerndal M, Forsberg K, Sena ES, et al. A systematic review and meta-analysis of erythropoietin in experimental stroke. *J Cereb Blood Flow Metab* 2010;30(5):961-8. doi: 10.1038/jcbfm.2009.267 [published Online First: 2009/12/30]
16. Lino de Oliveira C. A protocol for systematic review and meta-analysis of data from preclinical studies employing the forced swimming test, CAMARADES, 2016.
17. Lino de Oliveira C, Bahor Z, Currie G, et al. Search strategy for the protocol: A protocol for systematic review and meta-analysis of data from preclinical studies employing the forced swimming test. 26/02/2016 ed: CAMARADES, 2016.
18. de Vries RBM, Hooijmans CR, Langendam MW, et al. A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. *Evidence-based Preclinical Medicine* 2015;2(1):1-9. doi: 10.1002/ebm2.7
19. Sena ES, Currie GL, McCann SK, et al. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. *J Cereb Blood Flow Metab* 2014;34(5):737-42. doi: 10.1038/jcbfm.2014.28 [published Online First: 2014/02/19]