SYSTEMATIC REVIEW PROTOCOL FOR ANIMAL INTERVENTION STUDIES

FORMAT BY SYRCLE (WWW.SYRCLE.NL)
VERSION 2.0 (DECEMBER 2014)

Item #	Section/Subsection/Item	Description	Check for approval
	A. General		
1.	Title of the review	Sensitivity of mouse behavioural tests of anxiety to anxiolytic drugs approved for treatment of anxiety in humans.	
2.	Authors (names, affiliations, contributions)	Marianna Rosso Hanno Würbel Bernhard Voelkl - Animal Welfare Division, Veterinary Public Health Institute, University of Bern, Switzerland.	
3.	Other contributors (names, affiliations, contributions)		
4.	Contact person + e-mail address	Marianna Rosso (marianna.rosso@vetsuisse.unibe.ch)	
5.	Funding sources/sponsors	SNF Grant 310030_179254	
6.	Conflicts of interest	None	
7.	Date and location of protocol registration		
8.	Registration number (if applicable)		
9.	Stage of review at time of registration	Selection phase, Drug Selection: complete Selection phase, Behavioural Tests Selection: complete Selection phase, Pool screening: complete Selection phase, Title and Abstract screening: complete	
	B. Objectives		
	Background		
10.	What is already known about this disease/model/intervention? Why is it important to do this review?	In the drug development process, animal models are a key component of the preclinical phase, where drugs undergo laboratory testing to answer basic questions about safety and efficacy before being tested on humans. A growing body of evidence shows poor translation from animal research into human outcomes. This generates scientific, medical, economic and ethical concerns, including concerns for patients at risk and animal welfare. Therefore, action is needed to mitigate these concerns and improve translation from animals to humans. In terms of animal welfare, the implications of poor translation are twofold. On one hand, if animal tests are not predictive of human outcomes, the use of animals for such tests lacks both scientific and ethical justification. On the other hand, some animal tests, such as test of anxiety or pain are also important in view of assessing animal welfare. Therefore, we need animal tests that are valid in terms of the animal outcomes and predictive for the clinical (or human) outcomes. We hypothesise that part of the translational failure is a	

		lack of construct validity of the animal tests. For example, to study anxiety in mice, there are many different behavioural tests, each with a number of different outcome measures, not all of which may have sufficient construct validity. Poor construct validity is likely associated with poor predictive validity, which depends on the sensitivity and specificity of outcome measures to clinically effective treatments. Here, we will assess the sensitivity of outcome measures of mouse behavioural tests of anxiety to anxiolytic drugs with demonstrated efficacy in humans, in view of further studying the construct validity of mouse behavioural tests of anxiety and its implications for translational failure.	
	Research question		
11.	Specify the disease/health problem of interest	Anxiety.	
12.	Specify the population/species studied	Laboratory mice.	
13.	Specify the intervention/exposure	Administration of anxiolytic drugs (listed in Supplementary Table 1).	
14.	Specify the control population	Control group (receiving, for instance, vehicle) for each study.	
15.	Specify the outcome measures	Size of treatment effect (differences between control group and treatment group) on outcome measures of each behavioural test as indicated in the Supplementary Table 3.	
16.	State your research question (based on items 11-15)	Are mouse behavioural tests of anxiety sensitive to anxiolytic drugs approved for treatment of anxiety in humans?	
	C. Methods		
	Search and study identification		
17.	Identify literature databases to search (*e.g.* Pubmed, Embase, Web of science)	☐MEDLINE via PubMed ☐Web of Science ☐SCOPUS ☐EMBASE ☐Other, namely: ☐Specific journal(s), namely:	
18.	Define electronic search strategies (*e.g.* use the step by step search guide[15] and animal search filters[20, 21])	When available, please add a supplementary file containing your search strategy: [see last parts]	
19.	Identify other sources for study identification	☐Reference lists of included studies ☐Books ☐Reference lists of relevant reviews ☐Conference proceedings, namely: ☐Contacting authors/ organisations, namely: ☐Other, namely:	
20.	Define search strategy for these other sources	NA	
	Study selection		
21.	Define screening phases (*e.g.* pre-	1. Pool search result in reference management	

	screening based on title/abstract, full text screening, both)	program (Citavi) and remove duplicates. 2. Screen of Title and Abstract according to criteria below. 3. Full-text screening of papers passing step 2.	
22.	Specify (a) the number of reviewers per screening phase and (b) how discrepancies will be resolved	a) One reviewer per pool search result and screen of title and abstract; two reviewers for the full-text screening. b) Discrepancies will be solved by discussion.	
	Define all inclusion and exclusion criteria based on:		
23.	Type of study (design)	Inclusion criteria: Experimental, in vivo studies. Exclusion criteria: Others.	
24.	Type of animals/population (*e.g.* age, gender, disease model)	Inclusion criteria: Mice. Exclusion criteria: Other species.	
25.	Type of intervention (*e.g.* dosage, timing, frequency)	Inclusion criteria: administration of anxiolytic drugs listed in Supplementary Table 1. Exclusion criteria: studies using other or no drugs.	
26.	Outcome measures	Inclusion criteria: - Exclusion criteria: papers in which data is not shown neither in the main text, nor in the supplementary material. Cases where mean is not reported, will be excluded. Additionally, cases where neither standard deviation nor standard error are reported, are excluded. If the standard error is reported, cases where the sample size is not reported are excluded.	
27.	Language restrictions	Inclusion criteria: English. Exclusion criteria: Others.	
28.	Publication date restrictions	Inclusion criteria: - Exclusion criteria: none	
29.	Other	Inclusion criteria: - Exclusion criteria: none	
30.	Sort and prioritize your exclusion criteria per selection phase	Selection phase: Drug Selection 1. Inclusion criteria: drugs commonly prescribed in humans for the treatment of anxiety[1,2,3] 2. Exclusion criteria: - Selection phase: Behavioural Test Selection After a comprehensive list of behavioural tests elaborated to assess anxiety in mice was done (Supplementary Table 2), each test was part of a preliminary research in the PubMed database, in combination with the list of drugs obtained in the previous selection step, and the species (detailed research in Appendix 1). 1. Inclusion criteria: each test that yielded more than 10 papers was included for further analysis. 2. Exclusion criteria: each test that yielded 10 papers or less was excluded from further analysis. Selection phase: Pool screening 1. Inclusion criteria: - 2. Exclusion criteria: exclude duplicates.	

		Selection phase: Title and Abstract screening 1. Inclusion criteria: papers for which it is not possible to identify, through screening of abstract and/or title, criteria listed in points 23-29. 2. Exclusion criteria: papers for which it is possible to identify, through the abstract and/or the title, that do not meet the inclusion criteria, or meet exclusion criteria, listed in points 23-29. Selection phase: Full text screening 1. Inclusion criteria: see full list in points 23-29. 2. Exclusion criteria: papers for which it is not possible to retrieve data regarding outcome measures indicated in point 35, neither from the main text nor from the supplementary material.	
	Study characteristics to be extracted (for assessment of external validity, reporting quality)		
31.	Study ID (*e.g.* authors, year)	• Authors. • Title. • Year. • Journal.	
32.	Study design characteristics (*e.g.* experimental groups, number of animals)	• Sample size for control and treatment group from which data is to be extracted.	
33.	Animal model characteristics (*e.g.* species, gender, disease induction)	• Sex. • Strain. • Transgenic ID. • Treatment of stress (including CUMS) or social defeat. In case the animals are exposed to any further treatment or conditioning (for instance chronic stress, diabetes, pregnancy, etc.) the following rules will apply: **Study**: both *Control* (Non-Conditioned, NC) vs. *Drug treatment* (NC) and *Control* (Conditioned, C) *vs*. *Drug treatment* (C) are reported: → only data from Control (NC) vs. Drug treatment (NC) will be extracted. **Study**: only *Control* (Conditioned) vs. *Drug treatment* (Conditioned) is reported: → all data from *Control* (C) vs. *Drug treatment* (C) will be extracted. **Study**: *Control* (NC) vs. *Drug treatment* (C) (or vice-versa) is reported: → paper will be excluded.	
34.	Intervention characteristics (*e.g.* intervention, timing, duration)	• Drug(s) administered, If more than one drug is administered in the study, data will be extracted for each drug, even in	

		reference to the same control group. If drugs in Supplementary table 1 are used only in combination with other compounds (and therefore there is no information regarding drug vs. control), the study will be excluded.
		
		• Dosage (mg/kg),
		If the dosage is not reported in mg/kg, and it is not possible to convert the measure reported in mg/kg, the paper will be excluded.
		If more than one drug dosage is administered in the study, the following criteria will be pursued:
		  • If an uneven number of dosages was used in the study, the median dosage will be used.
		  • If two dosages are equally close to the median, one of the two will be randomly selected (see Appendix 2, randomization method).
		• Time of administration before behavioural test started. If the drug is administered chronically, this will be recorded in this section. If, in the same study, a drug is administered both chronically and acutely (in a single dose), only measures from single doses administration will be considered.
		• Route of administration.
		• Test duration,
		If more than one test duration is provided in the study, the following criteria will be pursued:
		  • If an uneven number of durations was used in the study, the median duration will be used.
		  • If two test durations are equally close to the median, one of the two will be randomly selected. (see Appendix 2, randomization method).
		
		Further decision criteria:
		• If a treatment is repeated both in home cage and in test apparatus, only the outcome measures from the test apparatus will be considered.
		• If the treatment is administered to mothers, and the offspring is subjected to the behavioural test, this will be excluded from further analysis.	

35.	Outcome measures	• Behavioural test(s) used. Outcome measures for each behavioural test are outlined in Supplementary Table 3, and are determined based on the construct of each behavioural test. For each, we will record mean values, sample size and standard deviation or standard error of the mean, for both treatment and	

		control group. In case of repeated measures on the same animals, only one outcome measure will be selected according to the following criteria: • In case of a before/after an additional intervention (e.g. surgery) study, the "before" condition will be selected, as a more appropriate indication of baseline state. • In case of studies repeated at different ages, the measures closer to adulthood (in mice, 3 to 6 months of age) will be selected. If two measures are equally distant from the mean of 4.5 months of age (or 18 weeks), one of the two will be randomly chosen (see Appendix 2, randomization method). For any additional cases, a decision will be made by discussion between the two reviewers. The decision and the reasoning will be subsequently reported.	
36.	Other (*e.g.* drop-outs)	-	
	Assessment risk of bias (internal validity) or study quality		
37.	Specify (a) the number of reviewers assessing the risk of bias/study quality in each study and (b) how discrepancies will be resolved	a) Two reviewers assessing risk of bias b) Discrepancies will be solved by discussion	
38.	Define criteria to assess (a) the internal validity of included studies (*e.g.* selection, performance, detection and attrition bias) and/or (b) other study quality measures (*e.g.* reporting quality, power)	☐By use of SYRCLE's Risk of Bias tool[4] ☐By use of SYRCLE's Risk of Bias tool, adapted as follows: ☐By use of CAMARADES' study quality checklist, e.g [22] ☐By use of CAMARADES' study quality checklist, adapted as follows: ☐Other criteria, namely: adaptation of CONSORT statement (consort-statement.org) according to the following criteria: Due to a large number of papers obtained from the search, a random sub-sample of 100 papers will be drawn from the total sample of papers and a detailed analysis of study quality will be conducted on this sub-sample.	
	Collection of outcome data		
39.	For each outcome measure, define the type of data to be extracted (*e.g.* continuous/dichotomous, unit of measurement)	Outcome measure will be recorded as provided by the authors. They may include both continuous (e.g. time duration, distance travel, etc.) or discrete (e.g. number of squares crossed, etc.) types of data.	
40.	Methods for data extraction/retrieval (*e.g.* first extraction from graphs using a digital screen ruler, then contacting authors)	1. Data extraction from tables and text. 2. If no numerical data is available, digitalization of tables and/or graphs (Automeris, https://apps.automeris.io/wpd) will be used to extract data.	
41.	Specify (a) the number of reviewers extracting data and (b) how	a) Two reviewers for full-text screen and data extraction	

	discrepancies will be resolved	b) Discrepancies will be solved by discussion	
	Data analysis/synthesis		
42.	Specify (per outcome measure) how you are planning to combine/compare the data (*e.g.* descriptive summary, meta-analysis)	The different outcome measures will be grouped according to measure type and analysed separately.	
43.	Specify (per outcome measure) how it will be decided whether a meta-analysis will be performed	A meta-analysis will be performed for each measure type unless insufficient data is retrieved from the papers scan.	
	If a meta-analysis seems feasible/sensible, specify (for each outcome measure):		
44.	The effect measure to be used (*e.g.* mean difference, standardized mean difference, risk ratio, odds ratio)	According to the type of data found in the papers, it will be either standardized mean difference or odds ratio.	
45.	The statistical model of analysis (*e.g.* random or fixed effects model)	Random effect model for the meta-analysis.	
46.	The statistical methods to assess heterogeneity (*e.g.* I^2, Q)	I^2	
47.	Which study characteristics will be examined as potential source of heterogeneity (subgroup analysis)	The following variable will be sub-grouped: -sex -strain -health condition (stress or social defeat)	
48.	Any sensitivity analyses you propose to perform	NA	
49.	Other details meta-analysis (*e.g.* correction for multiple testing, correction for multiple use of control group)	NA	
50.	The method for assessment of publication bias	A funnel plot will be drawn.	
Final approval by (names, affiliations):		SYRCLE	Date: Nov 2019

Supplementary Table 1: list of drugs included in the systematic review

Alprazolam
Amitriptyline
Buspirone
Chlordiazepoxide
Citalopram
Clomipramine
Clonazepam
Clorazepate
Desipramine
Diazepam
Doxepin
Duloxetine

Escitalopram
Fluoxetine
Flurazepam
Fluvoxamine
Hydroxyzine
Imipramine
Lorazepam
Maprotiline
Mirtazapine
Nortriptyline
Oxazepam

Paroxetine
Protriptyline
Sertraline
Temazepam
Trazodone
Triazolam
Trimipramine
Venlafaxine

Supplementary Table 2: Identified behavioural tests assessing anxiety[4,5,6,7] (Behavioural tests), number of papers yielded by the search on PubMed as explained in the text, subsequent inclusion in following steps of the review.

Behavioural tests	**Number of papers**	**Include (Yes/No)**
Elevated Plus Maze	330	Yes
T maze	17	Yes
Zero maze	14	Yes
Y maze	22	Yes
Elevated open platform	2	No
Open field	297	Yes
Light dark box	71	Yes
Emergence test	1	No
Social interaction test	72	Yes
Graded anxiety	1	No
Successive alleys	0	No
Triple test	2	No
Defensive marble burying	1	No
Shock probe burying	2	No
Ultrasonic vocalization	3	No
Staircase test	16	Yes
Suok ropewalking	1	No
Seed finding test	0	No
Anxiety/Defense test battery (MDTB)	10	No
Holeboard test	39	Yes
Geller-Seifter test	3	No
Vogel (punished drinking) test	17	Yes
Four plate test	28	Yes
Novelty suppressed feeding	49	Yes
Conditioned emotional response	5	No
Fear-potentiated startle	6	No
Active/passive avoidance	0	No
Predator odor test	3	No

Supplementary Table 3: Outcome measures for each behavioural test

i. OF, open field
   i. **sqs**: number of squared crossed.
   ii. **rear**: number of rearings.
   iii. **dist**: distance travelled.
   iv. **per_cent:** % of time spent in the center (as defined by the authors).
   v. **time_cent**: time (sec) spent in the center (as defined by the authors).

ii. EPM, elevated plus maze
   vi. **poa**: % of time spent in the open arms.
   vii. **toa**: time (sec) spent in the open arms.
   viii. **eoa**: Number of entries in the open arms.
   ix. **eca**: Number of entries in the closed arms.

iii. EZM, elevated zero maze
   x. **poc**: % of time spent in the open compartment.
   xi. **toc**: time (sec) spent in the open arms.
   xii. **eoc**: Number of entries in the open compartment.
   xiii. **ecc**: Number of entries in the closed compartment.

iv. HBT, holeboard test
   xiv. **hd**: Number of head dips.

v. LDB, light dark box
   xv. **light**: % of time spent in the light compartment.
   xvi. **t_light**: time (sec) spent in the light compartment.
   xvii. **trans**: number of transitions between the two compartments.

vi. SI, social interaction test

xviii. **time**: Time (sec) spent in social interaction.

vii. STC, staircase test

xix. **stps**: Number of steps climed.

xx. **rrs**: Number of rearings.

viii. VT, vogel conflict test

xxi. **shck**: Number of shocks accepted or received.

xxii. **dbs**: Number of drinking bouts.

ix. NSF, novelty suppressed feeding

xxiii. **lat**: Latency to eat (sec).

x. FPT, four-plate test

xxiv. **cross**: Number of punished crossings.

Appendix 1:

Research Class	Research string
1. Species	("mice"[All Fields] OR "Mice"[MeSH])
2. Boolean operator	AND
3. Behavioural tests	"*behavioural test according to Supplementary Table 2*"[All fields]
4. Boolean Operator	AND
5. Drugs	("Alprazolam"[Mesh] OR "Buspirone"[Mesh] OR Chlordiazepoxide"[Mesh] OR "Clorazepate Dipotassium"[Mesh] OR "Diazepam"[Mesh] OR "Doxepin"[Mesh] OR "Duloxetine Hydrochloride"[Mesh] OR "Citalopram"[Mesh] OR "Fluoxetine"[Mesh] OR "Flurazepam"[Mesh] OR "Fluvoxamine"[Mesh] OR "Hydroxyzine"[Mesh] OR "Imipramine"[Mesh] OR "Lorazepam"[Mesh] OR "Maprotiline"[Mesh] OR "Mirtazapine"[Mesh] OR "Nortriptyline"[Mesh] OR "Oxazepam"[Mesh] OR "Paroxetine"[Mesh] OR "Protriptyline"[Mesh] OR "Sertraline"[Mesh] OR "Temazepam"[Mesh] OR "Trazodone"[Mesh] OR "Triazolam"[Mesh] OR "Trimipramine"[Mesh] OR "Venlafaxine Hydrochloride"[Mesh] OR "Amitriptyline"[Mesh])

PubMed research string:

("mice"[All Fields] OR "Mice"[MeSH]) AND ("elevated plus maze"[All fields] OR "t maze"[All fields] OR "y maze"[All fields] OR "zero maze"[All fields] OR "0 maze"[All fields] OR "open field"[All fields] OR "light dark box"[All fields] OR "light dark test"[All fields] OR "light-dark box"[All fields] OR "light-dark test"[All fields] OR "social interaction"[All fields] OR "staircase test"[All fields] OR "hole board test"[All fields] OR "holeboard test"[All fields] OR "vogel test"[All fields] OR "punished drinking"[All fields] OR "four plate test"[All fields] OR "four plate"[All fields] OR "hyponeophagia test"[All fields] OR "novelty suppressed feeding"[All fields] OR "novelty induced hypophagia"[All fields]) AND ("Alprazolam"[Mesh] OR "Buspirone"[Mesh] OR "Chlordiazepoxide"[Mesh] OR "Clorazepate Dipotassium"[Mesh] OR "Diazepam"[Mesh] OR "Doxepin"[Mesh] OR "Duloxetine Hydrochloride"[Mesh] OR "Citalopram"[Mesh] OR "Fluoxetine"[Mesh] OR "Flurazepam"[Mesh] OR "Fluvoxamine"[Mesh] OR "Hydroxyzine"[Mesh] OR "Imipramine"[Mesh] OR "Lorazepam"[Mesh] OR "Maprotiline"[Mesh] OR "Mirtazapine"[Mesh] OR "Nortriptyline"[Mesh] OR "Oxazepam"[Mesh] OR "Paroxetine"[Mesh] OR "Protriptyline"[Mesh] OR "Sertraline"[Mesh] OR "Temazepam"[Mesh] OR

"Trazodone"[Mesh] OR "Triazolam"[Mesh] OR "Trimipramine"[Mesh] OR "Venlafaxine Hydrochloride"[Mesh] OR "Amitriptyline"[Mesh])

Embase research string:
'mouse'/exp OR mouse AND ('elevated plus maze' OR 't maze' OR 'y maze' OR 'zero maze' OR '0 maze' OR 'open field' OR 'light dark box' OR 'light dark test' OR 'light-dark box' OR 'light-dark test' OR 'social interaction' OR 'staircase test' OR 'hole board test' OR 'holeboard test' OR 'vogel test' OR 'punished drinking' OR 'four plate test' OR 'four plate' OR 'hyponeophagia test' OR 'novelty suppressed feeding' OR 'novelty induced hypophagia') AND ('alprazolam' OR 'buspirone' OR 'chlordiazepoxide' OR 'clorazepate dipotassium' OR 'diazepam' OR 'doxepin' OR 'duloxetine hydrochloride' OR 'citalopram' OR 'fluoxetine' OR 'flurazepam' OR 'fluvoxamine' OR 'hydroxyzine' OR 'imipramine' OR 'lorazepam' OR 'maprotiline' OR 'mirtazapine' OR 'nortriptyline' OR 'oxazepam' OR 'paroxetine' OR 'protriptyline' OR 'sertraline' OR 'temazepam' OR 'trazodone' OR 'triazolam' OR 'trimipramine' OR 'venlafaxine hydrochloride' OR 'amitriptyline')

Appendix 2: the toss of a coin will be used as a randomization method.

References:

1. DrugBank (drugbank.ca);
2. FDA Drug Approval Databases (fda.gov);
3. Anxiety and Depression American Association (adaa.org).
4. Hånell, Anders; Marklund, Niklas (2014): Structured evaluation of rodent behavioural tests used in drug discovery research. In Frontiers in behavioural neuroscience 8, p. 252. DOI: 10.3389/fnbeh.2014.00252.
5. Kumar, Vijender; Bhat, Zulfiqar Ali; Kumar, Dinesh (2013): Animal models of anxiety: a comprehensive review. In Journal of pharmacological and toxicological methods 68 (2), pp. 175–183. DOI: 10.1016/j.vascn.2013.05.003.
6. Harro, Jaanus (2018): Animals, anxiety, and anxiety disorders: How to measure anxiety in rodents and why. In Behavioural Brain Research 352, pp. 81–93. DOI: 10.1016/j.bbr.2017.10.016.
7. Crawley, Jacqueline N. (2007): What's wrong with my mouse? Behavioural phenotyping of transgenic and knockout mice. 2nd ed. Hoboken N.J.: Wiley-Interscience.