Pharmacological Research 196 (2023) 106917

Contents lists available at ScienceDirect

Pharmacological Research

journal homepage: www.elsevier.com/locate/yphrs

A systematic mapping review of the evolution of the rat Forced Swim Test: Protocols and outcome parameters

Christiane Brandwein [a,*,1], Cathalijn H.C. Leenaars [b,1], Laura Becker [a], Natascha Pfeiffer [a], Ana-Maria Iorgu [a], Melissa Hahn [a], Gaia A. Vairani [a], Lars Lewejohann [c,d], André Bleich [b], Anne S. Mallien [a], Peter Gass [a]

[a] Research Group Animal Models in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany
[b] Institute for Laboratory Animal Science, Hannover Medical School, 30625 Hannover, Germany
[c] German Federal Institute for Risk Assessment (BfR), German Center for the Protection of Laboratory Animals (Bf3R), Max-Dohrn-Str. 8-10, 10589 Berlin, Germany
[d] Animal Behavior and Laboratory Animal Science, Institute of Animal Welfare, Freie Universität Berlin, Königsweg 67, 14163 Berlin, Germany

ARTICLE INFO

Keywords:
FST
Porsolt
Systematic review
Rats
Forced Swim Test

ABSTRACT

As depression is projected to become the leading mental disease burden globally by 2030, understanding the underlying pathology, as well as screening potential anti-depressants with a higher efficacy, faster onset of action, and/or fewer side-effects is essential. A commonly used test for screening novel antidepressants and studying depression-linked aspects in rodents is the Porsolt Forced Swim Test. The present systematic mappping review gives a comprehensive overview of the evolution and of the most prevalently used set-ups of this test in rats, including the choice of animals (strain, sex, and age), technical aspects of protocol and environment, as well as reported outcome measures. Additionally, we provide an accessible list of all existing publications, to support informed decision-making for procedural and technical aspects of the test, to thereby enhance reproducibility and comparability. This should further contribute to reducing the number of unnecessarily replicated experiments, and consequently, reduce the number of animals used in future.

1. Introduction

According to the WHO report 2011, depression is projected to become the leading mental disease burden globally by 2030 [1]. This is still a realistic projection, as in 1990, mental disorders were the 13th leading cause of disability-adjusted life-years, climbing the ladder to the 7th rank by 2019, with depression being the most prevalent among the mental disorders [2]. Additionally, already in its first year, the COVID pandemic subsequently lead to an additional 3-fold increase of the depression rate [3]. Current treatments for depression are not only associated with a wide range of adverse side effects, withdrawal symptoms, and a therapeutic time lag of weeks to months (which is a topic for long discussions in itself [4]), they also still have a limited efficacy [5–7], including complete treatment resistance (TRD) [8]. Furthermore, an increasing number of adolescents are now a part of the severely affected population [9], and even the drugs currently classified as most effective need to be reevaluated in terms of their effect on the developing brain and general functioning [10]. The discovery of novel antidepressants as well as the reanalysis of current approaches are therefore essential. Unfortunately, alternative methods to replace animal models for drug development are not yet at a stage where they could provide the necessary deeper insight into this neurobiological disorder [11]. Hence, the persisting need for animal models for this very heterogeneous illness is obvious; not only do they enable insights into the neurobiological mechanisms and the etiology of depression, but they also facilitate the discovery and development of novel, more effective treatments with potentially fewer side effects.

The most commonly used test for screening antidepressant drugs and depression-linked aspects in rodents is the Forced Swim Test (FST), as introduced by Porsolt et al. in 1977 for rats [12] and for mice [13]. The popularity of this test comes as no surprise, as Cryan et al. [14] classified it as one of the two only high-ranking tests in the categories of reliability,

* Correspondence to: Department of Psychiatry and Psychotherapy Research Group Animal Models in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany.
E-mail address: Christiane.brandwein@zi-mannheim.de (C. Brandwein).
[1] shared first authors.

https://doi.org/10.1016/j.phrs.2023.106917
Received 5 July 2023; Received in revised form 22 August 2023; Accepted 8 September 2023
Available online 9 September 2023

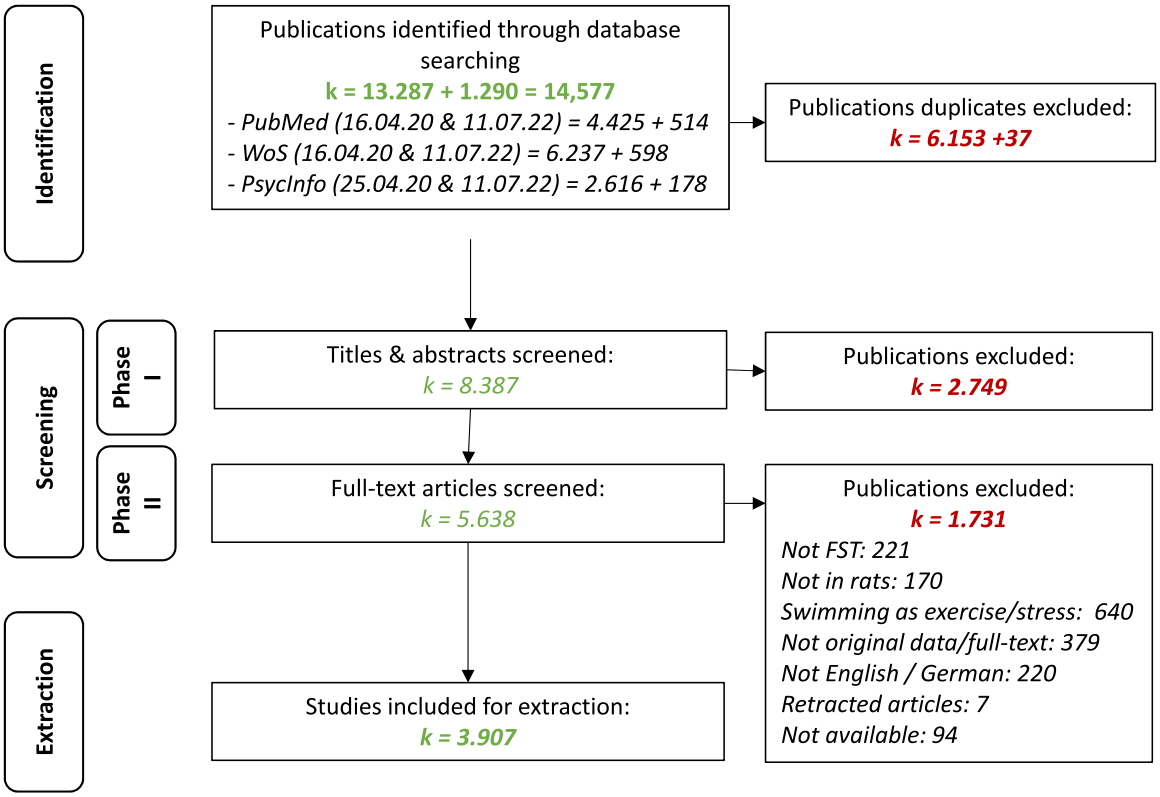

Fig. 1. Literature flow. Search results are split; the first number reflects the original search, the second number the update.

specificity, and ease of use. Even though the FST is said to have a high predictive validity, as it has been shown to reliably identify potent antidepressant drugs [15,16], its poor construct validity has increasingly become a topic of discussion [17,18]. This highlights a problematic aspect of the FST, as it appears to pick up on certain aspects of the pathophysiology of depression, but not on the full spectrum, as it is often simplified to do.

In the original protocol of the FST, the rat is placed into a cylinder with a width of about 20 cm, filled with 25 °C cold water, and its swimming behavior (immobility/floating versus active escape) is observed. While mice are typically only tested on a single day, the test is usually performed on two days in rats, with a 15 min acclimatization phase on day 1, before their behavior is recorded in a 5 min swim phase on day 2.

A healthy, untreated control animal will initially display swimming and "active escape" behavior (climbing and diving), but it will gradually start floating (described as "immobility"). The time spent immobile apart from the necessary movements to keep the nose above water, was originally considered the main read-out of the test, as it is decreased by a wide range of antidepressants [15]. Later, it was shown that different classes of antidepressants may evoke specific changes in other observable behaviors [19]. For example, norepinephrine-targeting antidepressants selectively increase vertical climbing behavior, while drugs influencing the serotonergic neurotransmission enhance horizontal swimming behavior. Including the differentiated observation of active escape behavior is commonly described as the "modified FST", although it only describes a modification of outcome measures.

However, the experimental design, animal strains used, and analyzed outcome measures vary widely, as the test has been implemented and modified by many laboratories, with the goal of optimizing it for specific research questions or their laboratory conditions. While such modifications can have obvious advantages within the respective laboratories, a great drawback is the overall incomparability of results, since even small adjustments of a parameter can have a large impact on the outcomes. Reproducibility of specific experiments is additionally hindered by poor reporting of the experimental details.

While several reviews have addressed the FST [15,20–22] and given an overview of a number of the above-mentioned aspects, none has done so using a comprehensive and systematic approach that provides a complete overview over the evolution as well as the currently most prevalently used set-ups of this test in rats. Therefore, we performed a systematic review (SR), to make existing data more accessible, support informed decision-making, and reduce the number of unnecessarily replicated experiments, and, with that, the number of animals used. We expect that new research approaches can be more easily refined (consistent with the 3Rs principle) by using the presented data, in terms of background strains, protocol set-ups and, in case of pharmacological tests, the most commonly used control substances. With our synopsis of data, we hope to contribute to increasing the value and reducing the unjustified use of animals in research studies [23].

Because of the sheer size of the literature on the FST in rats, and the resulting amount of data we extracted from the included studies, we decided to split up the results over multiple publications. This split is intended to increase legibility and coherence of the individual publications, and also to make this part of the data, which seems fundamentally relevant for the design of a potential experiment, accessible to the scientific community before we finish the full analyses of all data. Hence, the current publication focuses on which rat populations have been tested, and how the experimental design, set-up and readouts have evolved since the invention of the FST. In the supplements we share a full searchable spreadsheet of the included studies [24] to help scientists designing rat FST experiments easily retrieve the primary studies most relevant to their specific research questions.

1.1. *Tested animals*

We analyzed the choice of strains in the included papers, as the baseline behavior of different strains in the FST has been shown to vary.

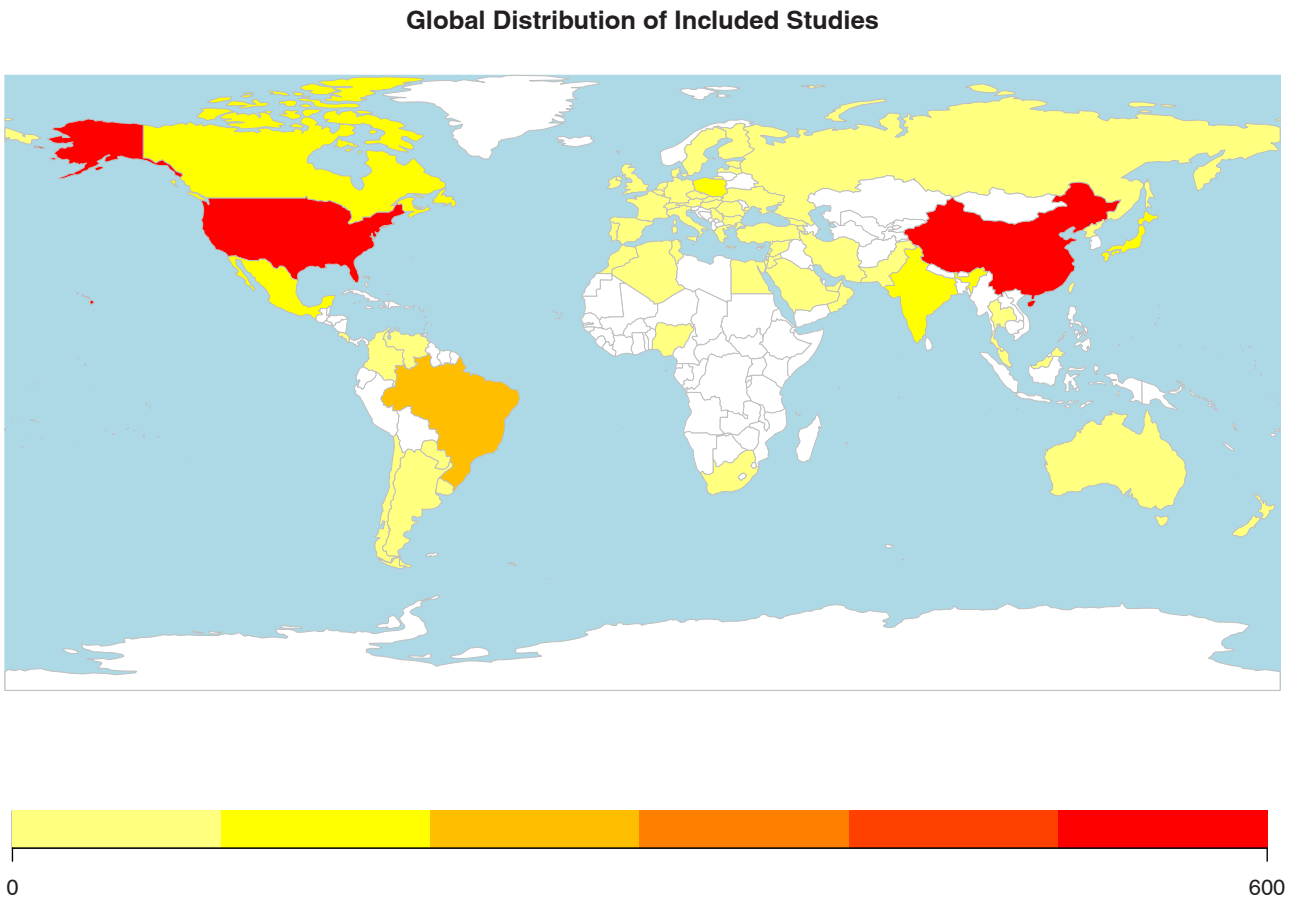

Fig. 2. World map showing the distribution of studies by country where the ethical approval was evaluated; white = no ethical approval reported within our dataset; light yellow = 1 – 200; dark red = 501 – 600 publications).

Table 1
Top 5 journals with the most publications included in this review.

Journal	Publications
Behavioural Brain Research	225
Psychopharmacology	130
Physiology & Behavior	128
Pharmacology, Biochemistry & Behavior	104
European Journal of Pharmacology	90
Other	3230

For example, the Flinders Sensitive Line was described to "hang" immediately when introduced into the water [25], which could create a ceiling effect in terms of immobility. Similarly, Wistar Kyoto appear to be extremely reactive to stress in comparison to other strains and display higher levels of baseline immobility in the FST compared to Wistar, Fischer 344, Lewis, and Sprague Dawley [26–28]. Additionally, strains appear to react differently to antidepressant treatment [28].

Another important factor to consider is the sex of the tested population. As in the past, scientists were cautious of introducing an extra source of variance due to physiological fluctuations caused by the estrous cycle and the sexually dimorphic Hypothalamic-Pituitary-Adrenal (HPA) axis, female animals were rarely included in experiments [22]. However, it has to be stressed that depression is twice as prevalent in the female population as in males [29]. Therefore, it is not appropriate to exclude females from predictive drug efficacy tests, since the response to antidepressants varies between the sexes [30]. Nor should they be excluded from research on the underlying pathophysiological mechanisms, if the aim is a generalizable applicability to clinical practice.

Choosing an adequate group size is a balancing act between statistical power and reducing animal numbers for welfare reasons. Yet, underpowered studies are likely to be contributing to the reproducibility crisis, hence we analyzed FST group sizes over time, to observe if awareness of this issue is increasing.

Not surprisingly, it has been shown that, with increasing age and weight, rats show more immobility in the FST [31]. Therefore, age and weight of tested animals can alter the observed outcomes, and we analyzed this parameter to show which age and weight ranges have been tested.

1.2. Technical aspects

In the original protocol, low water levels of 15–18 cm were used, where the animals could quite easily place their tail or feet on the bottom of the cylinder for support, which influenced the observed immobility time. It has been shown that immobility times decrease with deeper water (35 cm), while the corticosterone response remains unaffected by the different water depths [32,33]. Detke et al. [34] expanded these findings by discovering distinct differences in responses to various antidepressants: a broader behavioral repertoire of responses was observed in deeper water (swimming, struggling, climbing), which was later reproduced in other studies as well [14,35,36]. Specifically, SSRIs reduced immobility and increased horizontal swimming, whereas Norepinephrin reuptake inhibitors reduced immobility in favor of an increase in vertical activity, i.e. climbing [19,34]. Pinter et al. [37] revisited the water depths, among other parameters, and confirmed that immobility decreases significantly with higher water levels (at 30 cm and 45 cm), while struggling significantly increases, an effect that appears more profound in smaller than in larger animals. The

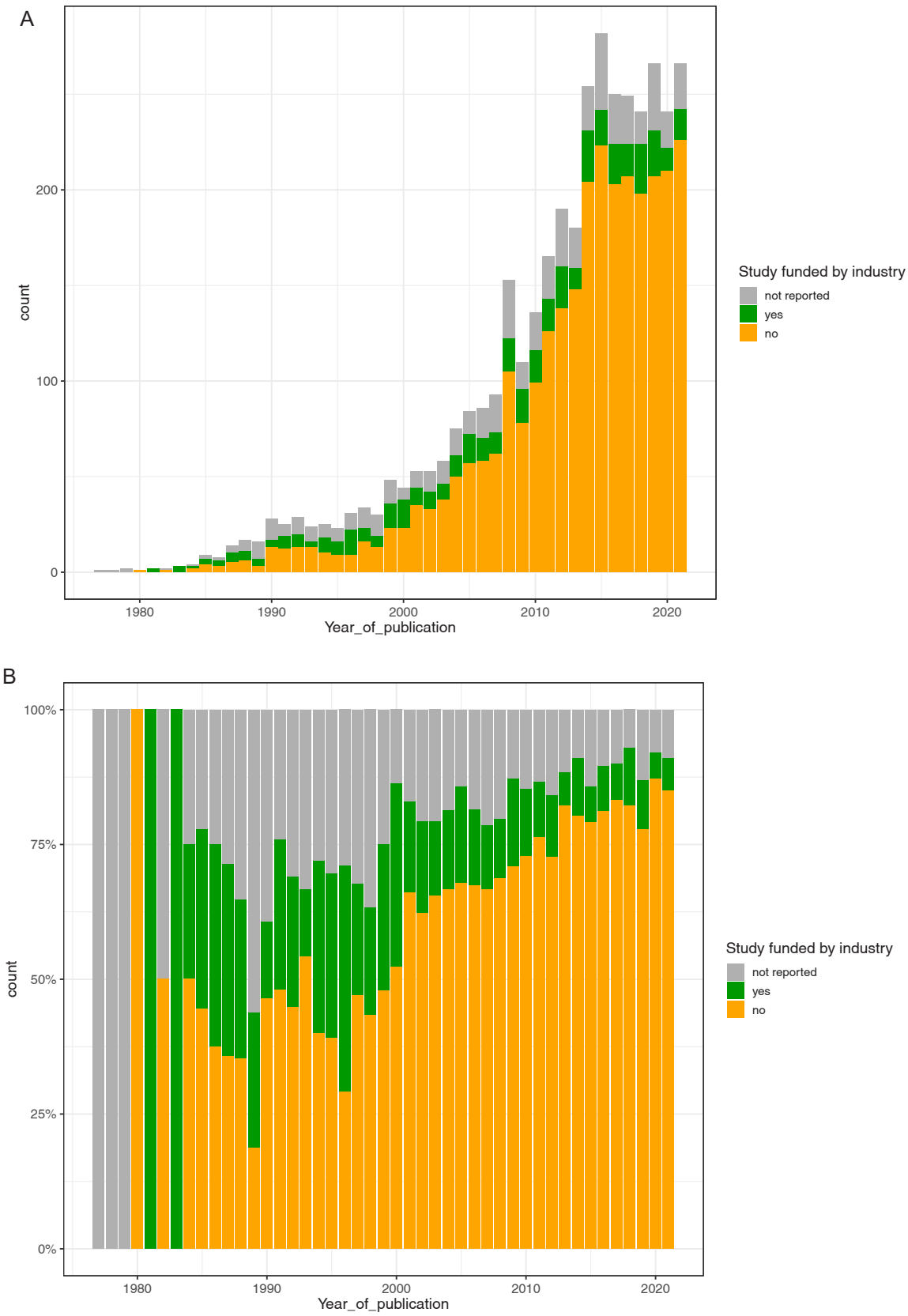

Fig. 3. Study funding in relation to published papers A) absolute numbers of publications B) percentages.

Table 2
Rat strains used for FST experiments.

Strain	Publications
Wistar	1629
Sprague Dawley (SD)	1621
Long Evans	107
Wistar Kyoto	75
Flinders Sensitive Line; Flinders Resistant Line	67
Other	343
not reported	1625

corticosterone levels seemed affected by the animal size more than by the water levels, again in line with the earlier findings of Abel et al. [32].

Another influential parameter to consider is the water temperature, as increased water temperature has been found to increase the floating time in the animals [37]. This may seem obvious, as colder water would expectedly be more aversive and therefore create a stronger urge to escape, besides increasing the need to generate heat through movement. As water temperature may not always be reflected upon when choosing the protocol for a specific research question, we also included analyses of this parameter in our review.

As the 1-day protocol is commonly used in mice, and the number of behavioral studies in mice is increasing, which could affect the relevance of the 1-day protocol in general, we were curious to see if there was an evolution in the application of this protocol in rats over time. Moreover, in some rat strains like the Wistar Kyoto, which are spontaneously immobile when set into the water, a "training day" seems unnecessary [38].

Bogdanova et al. reviewed that a larger diameter of the cylinder leads to a higher mobility [20]. In line with that idea, a larger diameter could lead to a more informative readout by reducing ceiling effects. Therefore, we also analyzed the chosen diameter within an experimental set-up.

In addition to affecting the natural activity of these nocturnal animals, the circadian rhythm also alters the effects of different antidepressants on immobility rates. Light, for example, possibly interferes with the dopaminergic receptors responsible for the anti-immobility effects induced by Desipramine [39]. Therefore, we also extracted the chosen diurnal cycle set-ups.

1.3. Outcome measures

We analyzed which outcome measures had been reported in the included studies, as the choice of the readout can significantly alter the interpretation of the observed results. For example, as mentioned above, including analyses of struggling, as well as the latency to immobility onset in mice [40], improves the accuracy of antidepressant efficacy [19]. Moreover, because the FST is a test that evaluates behavior which is heavily influenced by the locomotor abilities and tendencies of an animal in addition to possible stimulant effects of the tested antidepressant [41], it seems essential to include mobility as a possible confounding factor when interpreting immobility behavior.

1.4. Planned publications

While in this first foundational publication we focused on the essential parameters that need to be considered when designing an experiment (i.e. "*what* has been done before?"), future papers will describe a) how the aim and interpretation of this test have changed over time, (i.e. "*why* was it done?") and b) welfare aspects of the FST.

The interpretation of this test has been a topic of controversial discussions almost since its inception, as the observed immobility is either seen as behavioral despair or as passive coping strategy, which can be understood almost as polar opposites, albeit not mutually exclusive. Likewise, the welfare aspects of the FST have also been the focus of scientific and public debates [42], which have intensified in the past years, not lastly because the European Directive classified 'forced swim tests […] with exhaustion as the end-point' as *severe experiment* (Directive 2010/63/EU). The notable caveat of this classification, though, is that it solely addresses forced swimming with exhaustion as the endpoint. This is definitely more severe than the standard FST protocol, where the animals swim for a maximum of 15 min, while exhaustion is reached well after 120 min of swimming in untreated, unweighted controls [43,44]. We therefore find an evidence-based assessment of this test essential [45], as this classification is the foundation of ethics evaluations.

2. Methods

This systematic review is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards [46]. A systematic review protocol using the SYRCLE template [47] was developed to identify and review the relevant research on the evolution of the rat FST with respect to employed protocols, measured outcome parameters and research questions answered. This protocol was registered on 15 May, 2020 on SYRF (https://syrf.org.uk/protocols).

2.1. Search strategy

A comprehensive search strategy was developed to identify the peer-reviewed literature published in English or German between the conception of the Porsolt Forced Swim Test in rats in 1977 and the time of the performed searches. Search strings comprised title-abstract-keyword terms, and where applicable, thesaurus terms for the specific database (PubMed, Web of Science (WoS Core collection, BIOSIS, Social Science Citation Index), and PsychInfo via EBSCO). The full search strategy is published in the protocol, and the searches were performed on 16.04.2020 (PubMed and WoS) and on 25.04.2020 (PSYCInfo). The search was updated on 11.07.2022, to also retrieve newer publications, restricted to those from before 2022, to report only on complete years in all figures. Secondary reference searches ("snowballing") from reference lists of the included publications were not conducted, because of viability and limited expected additional value after the comprehensive searches in three large literature databases covering all relevant fields (medicine, biology, and psychology).

2.2. Screening for inclusion

After the removal of duplicates through the EndNote reference manager, a manual double-check as well as an automated triple-check through the Systematic Review Extraction platform Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia) were performed. Subsequently, titles and abstracts from all unique articles that were identified from the searches were reviewed in Covidence, by two out of eight trained, independent scientists (MH, CB, RA, CA, NP, LB, CL, FT).

Inclusion criteria were (i) publication was about the FST, (ii) test was performed in rats, (iii) swimming was not used as exercise or stressor, (iv) full-text publications with original data on swimming-related readout parameters, (v) in English or German. Other languages were excluded, as the review team would not have been able to accurately extract the experimental details from them. All abstracts which met the inclusion criteria then went into Phase 2 of the screening process in Covidence, the full-text screening. Each full text was screened by two out of seven trained, independent reviewers (MH, FT, CB, RA, NP, CA, LB).

Decision discrepancies within the two phases of the screening process were resolved by a third independent reviewer.

2.3. Data extraction

A template for data extraction was created in Covidence and was

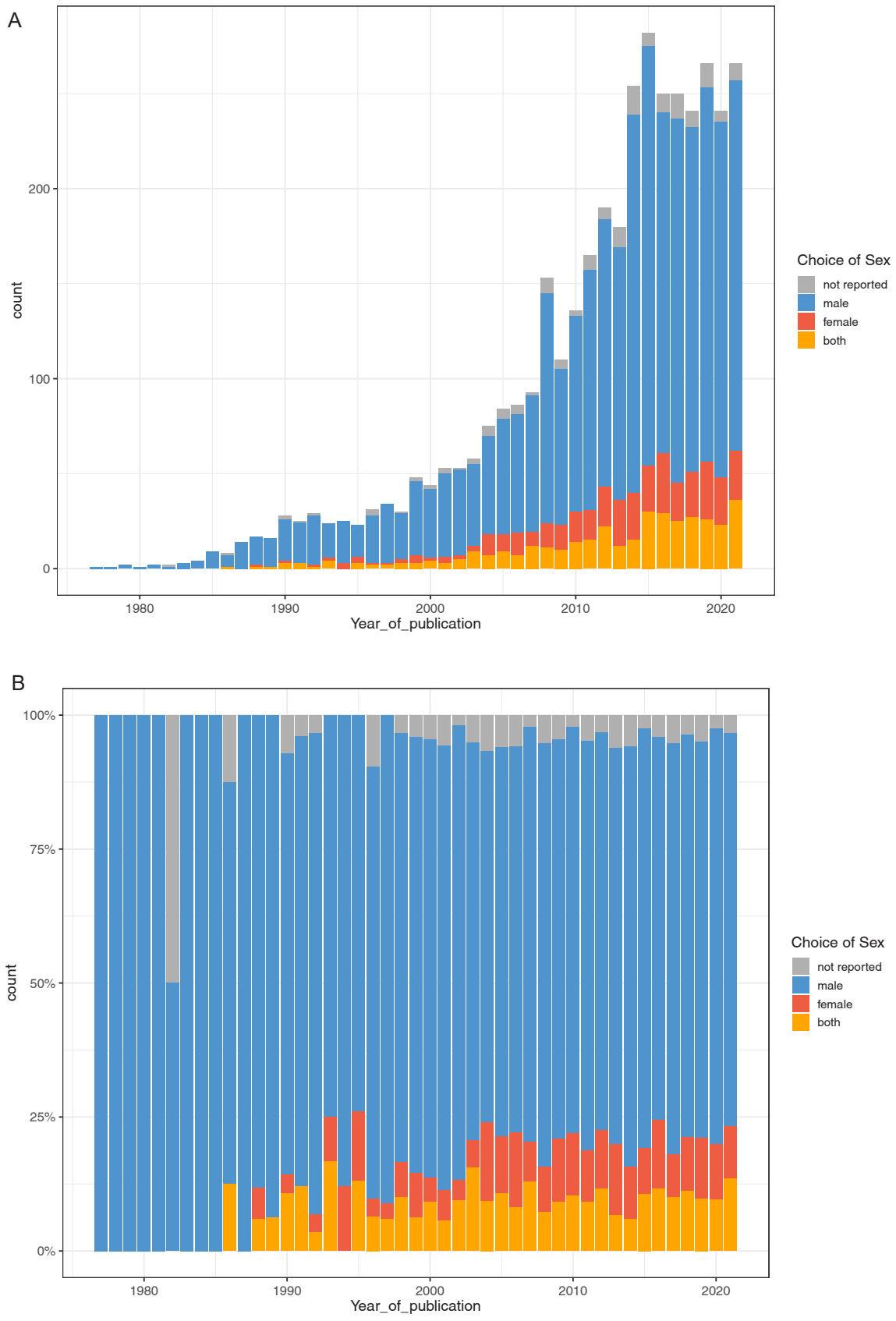

Fig. 4. Examined sex in relation to publications A) absolute numbers of publications B) percentages.

Table 3
Age and weight range of examined animals within the extracted publications; Qu. = Quartile, Min = minimum, Max = maximum, NA's = not reported as calculated from the factorial analysis *The publication, which had used 120 week old animals did not report their age at FST start, therefore the maximum age in this category was from another publication.

	Mean	Median	1st Qu.	3rd Qu.	Min	Max	NA's
Min AGE at project start (wks)	5.77	6.00	0.00	9.00	0.00	120.00	2129
Max AGE at project start (wks)	120.00	9.00	6.00	12.00	1.00	120.00	
Min AGE at FST start (wks)	11.17	9.00	7.86	12.00	1.00	87*	2981
Max AGE at FST start (wks)	15.07	12.00	11.00	16.00	1.50	120.00	
Min WEIGHT at project start (wks)	209.80	200.00	180.00	250.00	5.00	550.00	1270
Max WEIGHT at project start (wks)	264.70	250.00	220.00	300.00	7.00	1110.00	
Min WEIGHT at FST start (wks)	228.80	215.00	180.80	278.80	25.00	450.00	3708
Max WEIGHT at FST start (wks)	295.50	300.00	250.00	350.00	148.00	550.00	

Table 4
Top 10 antidepressants most frequently used in the rat FST.

Antidepressant	Publications
Fluoxetine	513
Imipramine	437
Desipramine	266
Ketamine	106
Citalopram	66
Amitriptyline	60
Clomipramine	57
Venlafaxine	53
Sertraline	42
Escitalopram	40
Other	340

used to extract data from the publications included in the review, as per protocol, by one reviewer. Training for data extraction of the participating reviewers encompassed a full check of the first 5 extractions to harmonize the process, then a double-check by an independent second reviewer (CB) of approximately 10% of the extractions, which was then reduced to 5% of random extraction double-checks once there were rarely any disagreements anymore.

2.4. Analysis

As mentioned above, this first publication focuses on experimental design and set-up. Analyses of these parameters is described here.

The "country of ethics evaluation" was marked as "not reported" if the ethics evaluation was not explicitly specified.

For the parameter "group size", if groups differed in size, the median of the range from the included study was analyzed. For simplicity and consistency, the parameter "housing" was categorized into the options "single" or "group", omitting the actual group sizes for this parameter. Diets were not considered as a "pharmacological treatment", unless the rats received something in addition to the diet, such as hormones etc. To ease the process, we extracted the parameters of weight and age as categorical ranges, but some extractors registered more detailed information. Therefore, descriptive statistics were determined from all extracted values, which were also the basis used for the diagrams. For the numerical age parameter, we omitted the categories "> 16 weeks", "prenatal", "juvenile" and "adult" from the calculations, and for the numerical weight parameter, we omitted "< 100 g" and "> 450 g". For the parameter "funding" we decided to use a stringent approach and already marked studies as "funded by industry" if solely the substance was donated by a company.

After completion of the extraction process and quality control, data were cleaned in Excel (homogenization of spelling, punctuation and capitalization, removal of spaces, calculation of medians or min and max values from ranges, translation of country names into ISO country codes, rechecking of incoherent data). Next, data were imported into R, version 4.1.0 "Camp Pontanezen" via RStudio [48], using the readxl package [49]. Summary statistics were calculated using the packages dplyr [50], crosstable [51], tidyr [52], and data.table [53]. Plots were created using ggplot2 [54], rworldmap [55] and tm [56,57].

As we were interested in comparing protocols, plots were created and simple statistics were calculated with the publication as the unit of analysis throughout, without correcting for the size of the studies described in them. Simple statistics were calculated with base R.

2.5. Risk of bias assessment

Most risk of bias tools focus on interventional studies [58], while we analyzed experimental design without a specific interventions. Therefore we did not perform a risk of bias analysis, comparable to the approach used by Van der Mierden et al. [59]. To get an indication of the quality of the included studies nonetheless, we assessed reporting of the following parameters as a proxy measure: (i) ethics approval, (ii) following international guidelines (NIH; ARRIVE; EU Directive; EEC Directive, which all include directions for bias minimization), (iii) study funding.

3. Results

Fig. 1 depicts the flow of the literature, from the search via title-abstract and full-text screening to inclusion.

3.1. Bibliographic details

3375 out of the total 3907 extracted studies reported, where the

Table 5
Water temperature, water depth, cylinder diameter, height and swim times as described in the extracted papers; Qu. = Quartile, Min = minimum, Max = maximum, NA's = not reported. *Note that the provided diameter in one of the papers was probably a typo (with 2.50 cm), therefore we excluded it and instead used the next-lowest value of 6 cm.

	Mean	Median	1st Qu.	3rd Qu.	Min	Max	NA's
Water temperature [°C]	24.37	25.00	24.00	25.00	18.00	36.00	735
Height of cylinder [cm]	48.60	46.00	40.00	52.00	15.00	150.00	719
Diameter* [cm]	24.17	20.00	20.00	30.00	6.00	152.00	732
Water depth [cm]	29.40	30.00	25.00	32.00	6.00	80.00	733
Swim time (accl day) [min]	14.44	15.00	15.00	15.00	1.00	60.00	1153
Swim time (test day) [min]	6.02	5.00	5.00	5.00	0.45	120.00	171

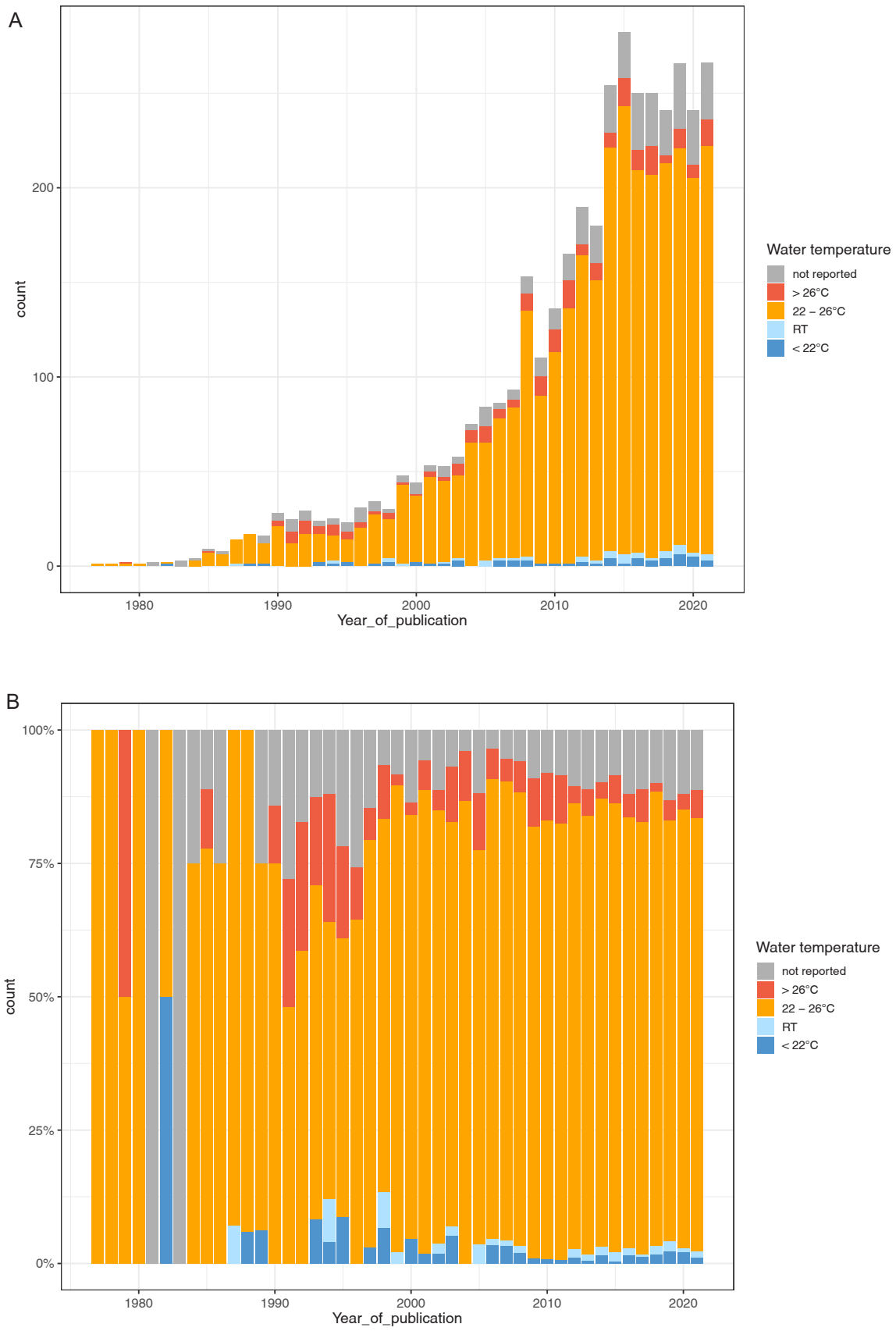

Fig. 5. Water temperature over time A) absolute numbers of publications B) percentages; RT = room temperature.

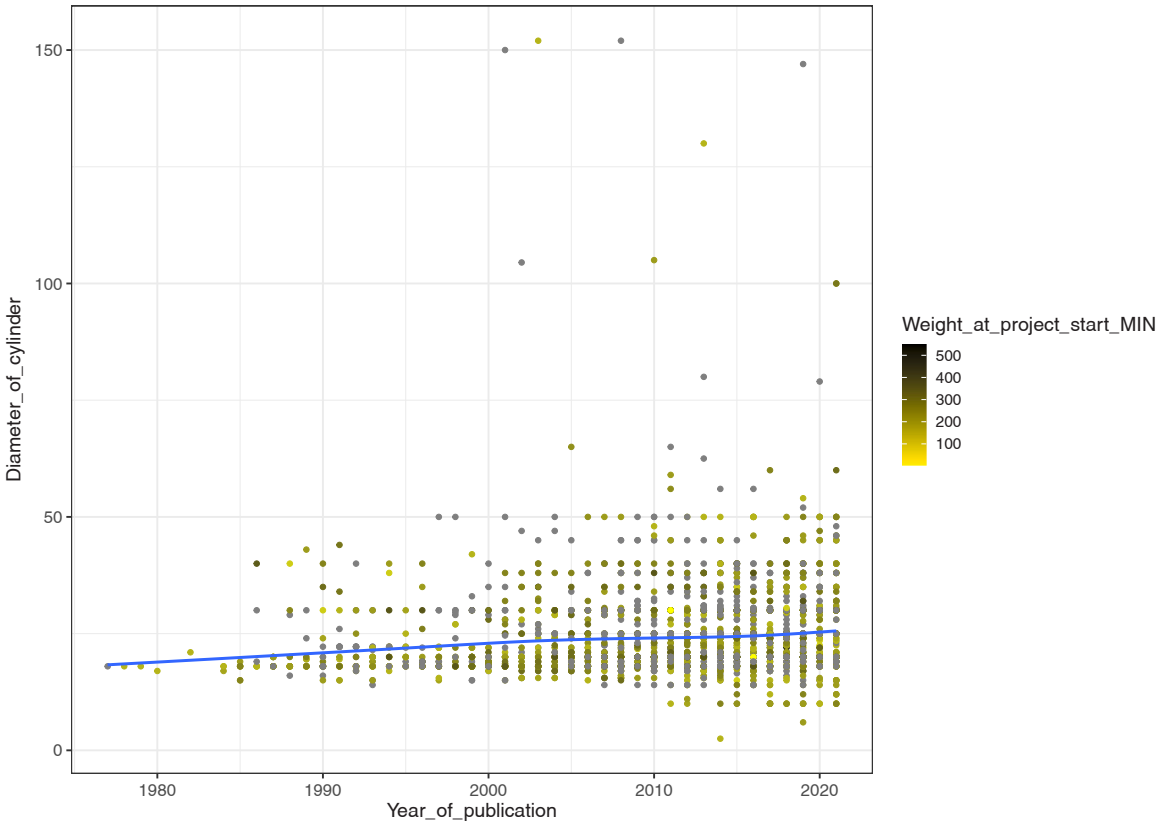

Fig. 6. Cylinder diameter in relation to publication date, with weight at project start depicted by different colors.

ethics of the experiment had been approved. This provides a more reliable representation of the location where the experiment was performed than the country of the first author would have. At the top of the list are the United States with 574 publications and China with 506, followed by Brazil with 290 publications in the analyzed timespan (Fig. 2). All extracted publications were in English, even though we would have included German as well.

The journal with the most publications in this review was "Behavioural Brain Research", with a tight run for number 2 and 3 by "Psychopharmacology" and "Physiology & Behavior" (Table 1). This ranking is an estimate, though, as this parameter was not fully harmonized in the final extraction (i.e. potentially sensitive to differences in free-text-writing).

Up to 2015, the number of FST publications increased (Fig. 3A), in line with overall publication numbers. There seems to be a stagnation from 2015 onwards. Compared to the overall number of FST studies, numbers of published studies supported by industrial funding seem to decrease over time (Fig. 3B). At the same time, the reporting on funding is apparently improving (Fig. 3B).

3.2. Tested animals

The rat strains used in the included studies are shown in Table 2. The most commonly used rat strains for the FST are Wistar and Sprague-Dawley. The three other strains that were used regularly were Long Evans, Wistar Kyoto, and Flinders. The "other" category comprised a wide variety of strains, e.g. Fischer, Albino, Lewis, Brown Norway, etc. Other strains, which are not mentioned in this table as they were each only reported in up to 20 publications, can easily be retrieved from the searchable spreadsheet [24].

Originally, research focused almost solely on male animals, as is visible in Fig. 4, most likely because of the historic perception that the estrous cycle introduced additional variance in the outcomes. Yet the use of females has increased since the first study reported to have successfully included both sexes in 1986 [60].

While they are highly relevant to the outcomes, the age and weight of the tested animals are rather poorly reported (Table 3), with only 2129 out of 3907 publications (54%) reporting the animal age. The animal weight at project start is described most often, as only 32% of data were not reported in this case, in comparison to 95% not reported weights at FST start (3708 out of 3907). In the studies that did report age during the FST, rat age ranged from 0 to 120 weeks. In the studies that described body weight during the FST, it ranged from 5 g to 1110 g (Table 3).

3.3. Control substances

"Control Substances" are clinically effective and approved antidepressants, categorized according to the list in Stahl's Essential Psychopharmacology [61], which were used as positive controls in the FST.

Table 4 summarizes the most frequently used antidepressants, with Fluoxetine (a selective serotonin reuptake inhibitor / SSRI) and Imipramine (a tricyclic antidepressant / TCA) appearing as the gold standards. Also at the top of the list are Desipramine (a TCA) and Ketamine (NMDAR antagonist), which was originally used as anesthetic and only discovered quite recently as having antidepressant effects with rapid onset even in treatment resistant patients [62].

3.4. Technical aspects

The selected water temperature ranged from 18 to 36 (Table 5), but largely adheres to the original protocol of Porsolt [12], where a temperature of 25 °C was used (Fig. 5).

The minimum diameter of 2.5 cm in Table 5 is most likely a typo within the included paper [63], and the actual minimum diameter probably was 6 cm [64]. The minimum water depth of 6 cm (Table 5) was repeated twice within the same lab, and is therefore probably not a

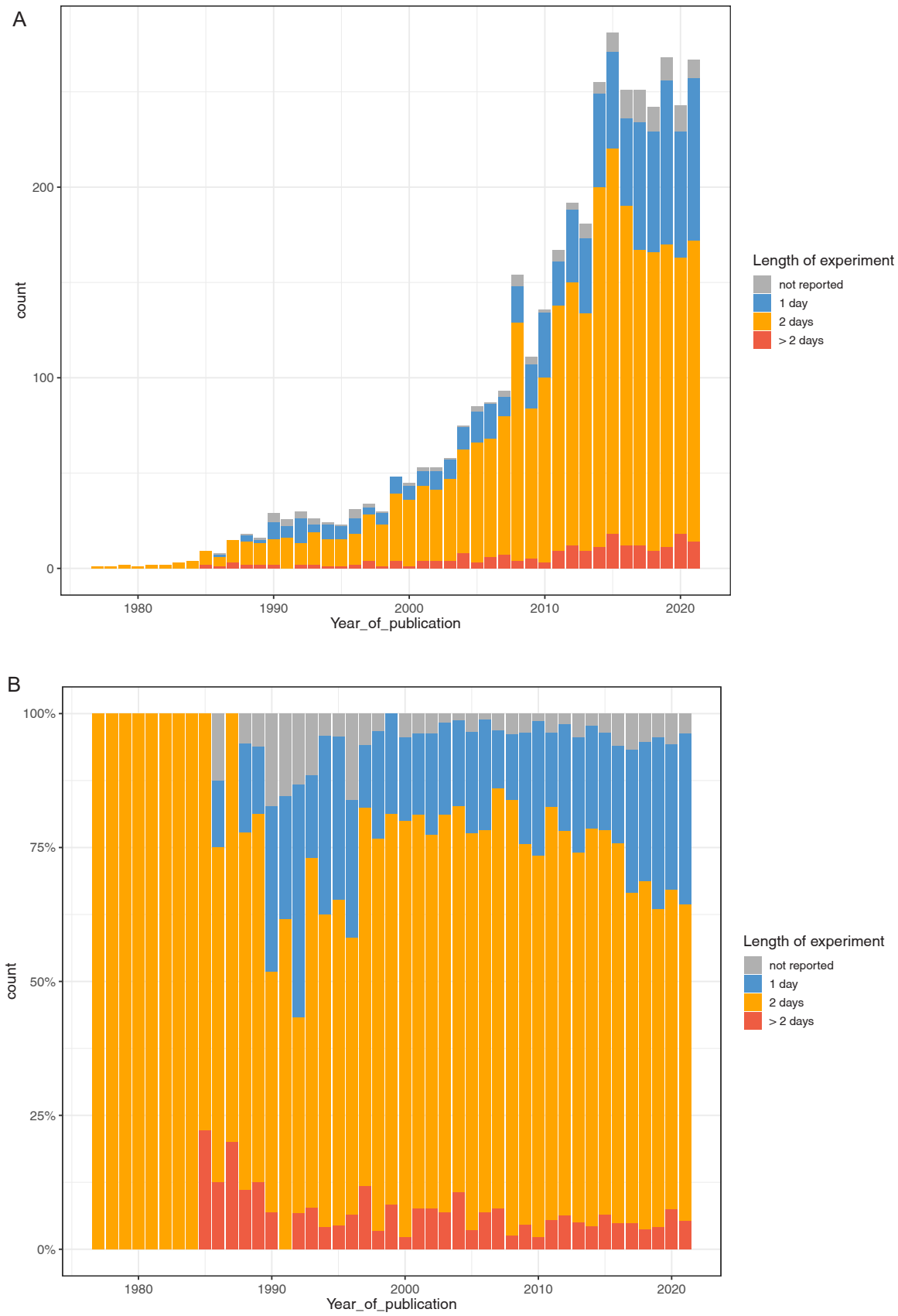

Fig. 7. Chosen length of experiment over time A) absolute numbers of publications B) percentages.

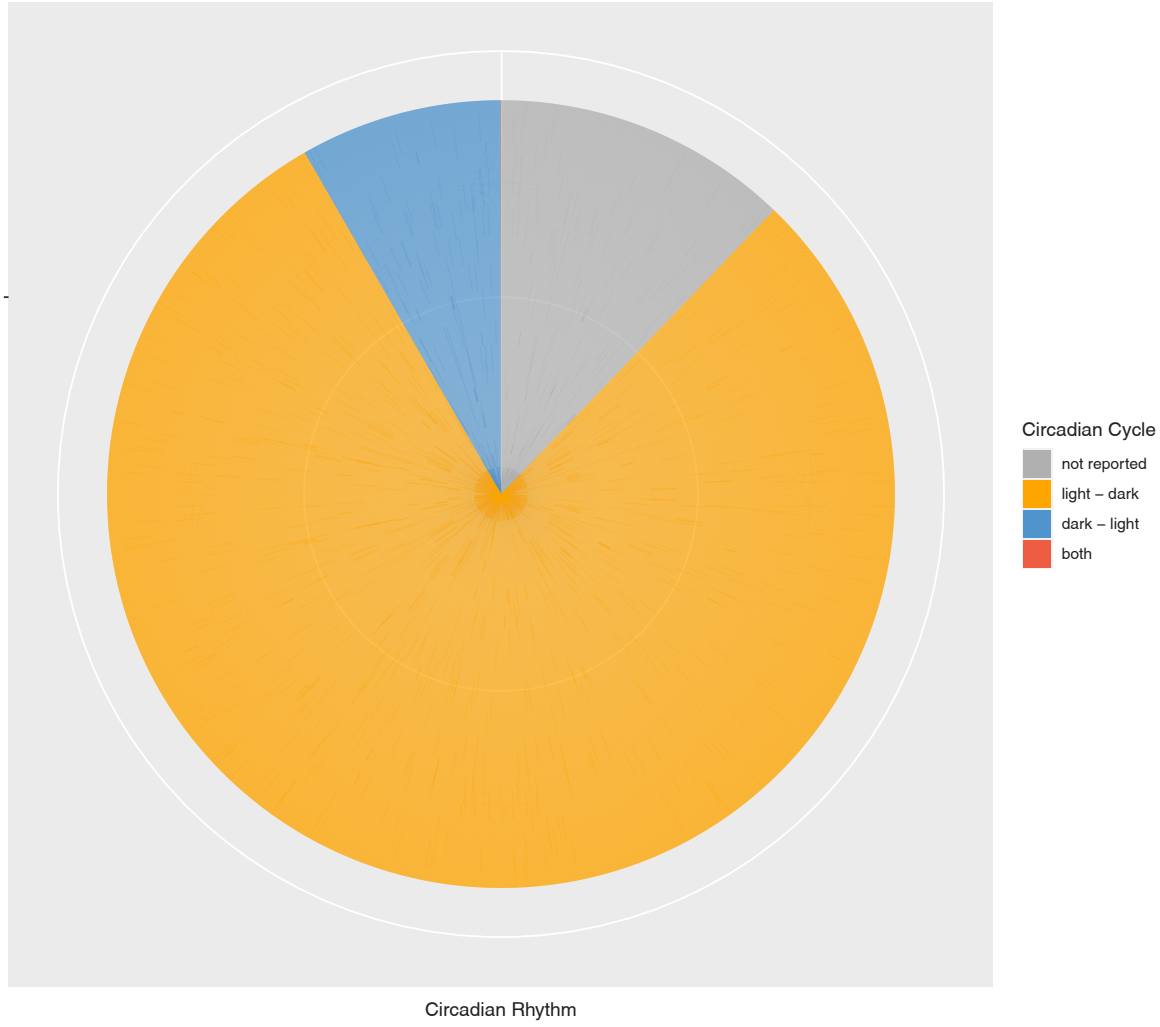

Fig. 8. Reported circadian cycles.

mistake [65,66]. The next lowest water depth was 9 cm [67].

As common sense would warrant the use of larger cylinders for larger animals, we analyzed the association of minimum weight at project start (the weight parameter with the least missing values, Table 3), diameter (Table 5) and publication date in the included studies. There was no obvious correlation, but in Fig. 6, one can see that, on average, the diameter has increased a little over time.

The swim times largely follow the original protocol as well, with a 15 min acclimatization session on the first day, followed by a 5 min test session 24 h after the initial exposure (Table 5). The outlier of a 60 min water exposure on the first day was only the set cutoff time, within which the animals were expected to float for a period of 120 s, after which they were promptly removed from the water [68]. However, according to the authors, none of the animals reached that limit. The number of NA's of the swim time on the acclimatization day includes 840 studies which used the one-day protocol and therefore did not have an acclimatization swim time.

Regarding the water exposure time on the testing day, most publications adhere to the original protocol as well, which observed the swimming behavior for 5 mins. The 4 publications, in which animals were only left to swim for 45 s were all performed in rat pups (please refer to the spreadsheet [24] for publications). The outlier in this set, which exposed the animals for a maximum of 120 mins was from 1988 [69], and again this was a cutoff timepoint, determined by the time it took the animal to sink, to be then immediately removed from the water. The next longest swim duration was a single publication that used 60 min water exposure to collect a series of microdialysis samples [70].

While the one day protocol is very common in mouse set-ups, it only started to be implemented in rats from 1986 [60] onwards. Since then, the relative use of the two-day protocol remained rather stable over the years nonetheless (Fig. 7 A and 7B). Reporting of the duration of the protocol was relatively complete, with only 171 (out of 3907 or 4%) studies not reporting this information.

Reporting of the light conditions of the housing situation was analyzed only at the level of standard versus reversed cycle. A normal light – dark cycle with testing during the light phase is clearly preferred by researchers reporting FST experiments, even though the animals are then tested during their inactive phase (Fig. 8). Only 88% of the studies reported this relevant parameter, though, and 323 studies out of the total 3907 described using a reversed dark-light cycle, with testing in the dark, active phase.

While the number of publications increased over time, the group sizes may appear to decrease, independent of the tested sex (Fig. 9). The four observed outliers which described large numbers of animals ($n > 50$) did not necessarily use more animals per experimental group, they mainly described complex experimental designs with multiple levels of grouping, where the exact n was not always clear at all levels. Group size was reported in 90% of the included studies.

3.5. Outcome measures

Various behavioral outcomes were analyzed for rats in the FST. The

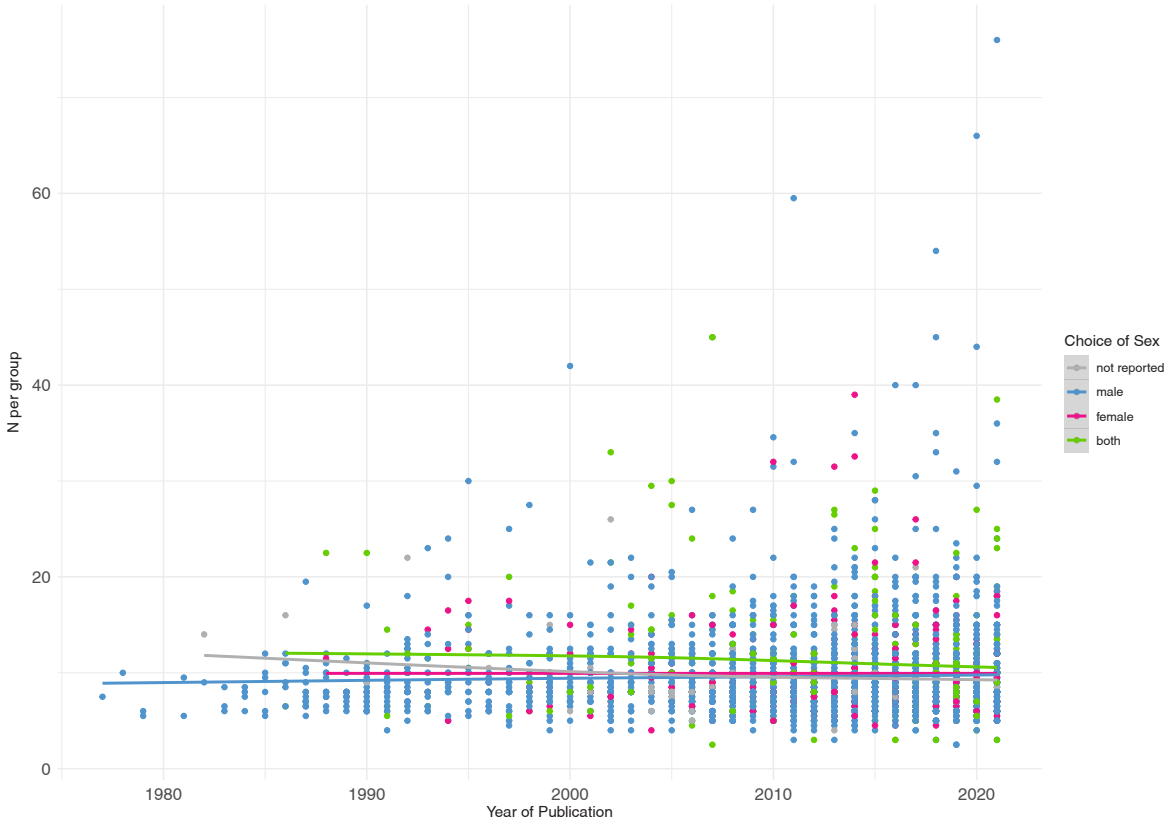

Fig. 9. Group size over time, with the sex depicted by different colors.

Table 6
Reported outcome parameters.

	Yes	No
Latency to first immobility	330	3577
Immobility	3796	111
Swimming	1434	2473
Climbing	1314	2593
Diving	176	3731
Locomotion	2196	1711

most commonly reported are listed in Table 6.

Even though Detke et al. already published the relevance of observing active escape attempts in addition to the conventional immobility parameter in 1995 [19], implementation of this outcome measure only appears to increase after 2002 (Fig. 10).

The latency at which the animal starts to float is a very important parameter in the mouse set-up of the FST as it has been shown to increase the predictive validity of the test for tricyclic antidepressants as well as norepinephrine reuptake inhibitors [21]. In the rat FST it is still less common, with only 330 out of 3907 studies reporting this outcome (Table 6, Fig. 11).

Overall, 2196 out of 3907 studies analyzed locomotion parallel to the FST, a trend that has grown over time (Fig. 12). We did not formally analyze how many publications statistically looked at the correlation between FST and locomotion readouts, or used locomotion measures to correct FST outcomes and prevent confounding, but in our subjective experience during data extraction this was rare.

4. Discussion

The aim of this first publication of our Systematic Mapping Review was to describe the technical evolution of the Porsolt Forced Swim Test, as well as the currently most prevalently used set-ups of this test, as a foundation for evidence-based decision-making for future experimental designs. This review is, to the best of our knowledge, the most comprehensive overview of rat FST publications to date. Because of the large number of included publications, however, data extraction was restricted to relatively straightforward measures. To make the data available to the scientific community, we postpone more in-depth analyses to future publications and, at this stage, shed light on the different modifications of the FST which have been introduced over the years. By making the collected data easily accessible, we support scientists in their informed decision-making process when planning future experiments, as the compiled data allow for simple retrieval of the specific literature relevant to their design.

The high predictive validity of the FST has convincingly been shown by Borsini et al.: 33 classic antidepressants performed well in the FST in comparison to the control substances across different laboratories, either in a single dose or with repeated treatments [15]. Therefore, the FST is and will most likely remain a standard for antidepressant screening, until an alternative approach to replace this animal model is found. Hence, this Systematic Review can be used to implement this test as efficiently as possible. This is crucial, as developing effective antidepressants is challenging; Trunnell et al. reviewed 109 therapeutics screened by the top 15 pharmaceutical companies and found that only seven of these compounds successfully made it through the FST. At this stage, however, none of these seven are marketed to treat human depression [71].

What appears as striking in the analysis, is that after an all-time high in publication counts on this test in 2015, there seems to be a stagnation in the use of the FST. This could be due to the increased awareness of welfare aspects over time, in addition to the increasingly obvious scientific discussion about readout interpretations, which gained traction again in 2015 [72].

With 86.4% of all publications, a large fraction of authors mentioned where the ethics of the experiments were approved. In a planned future

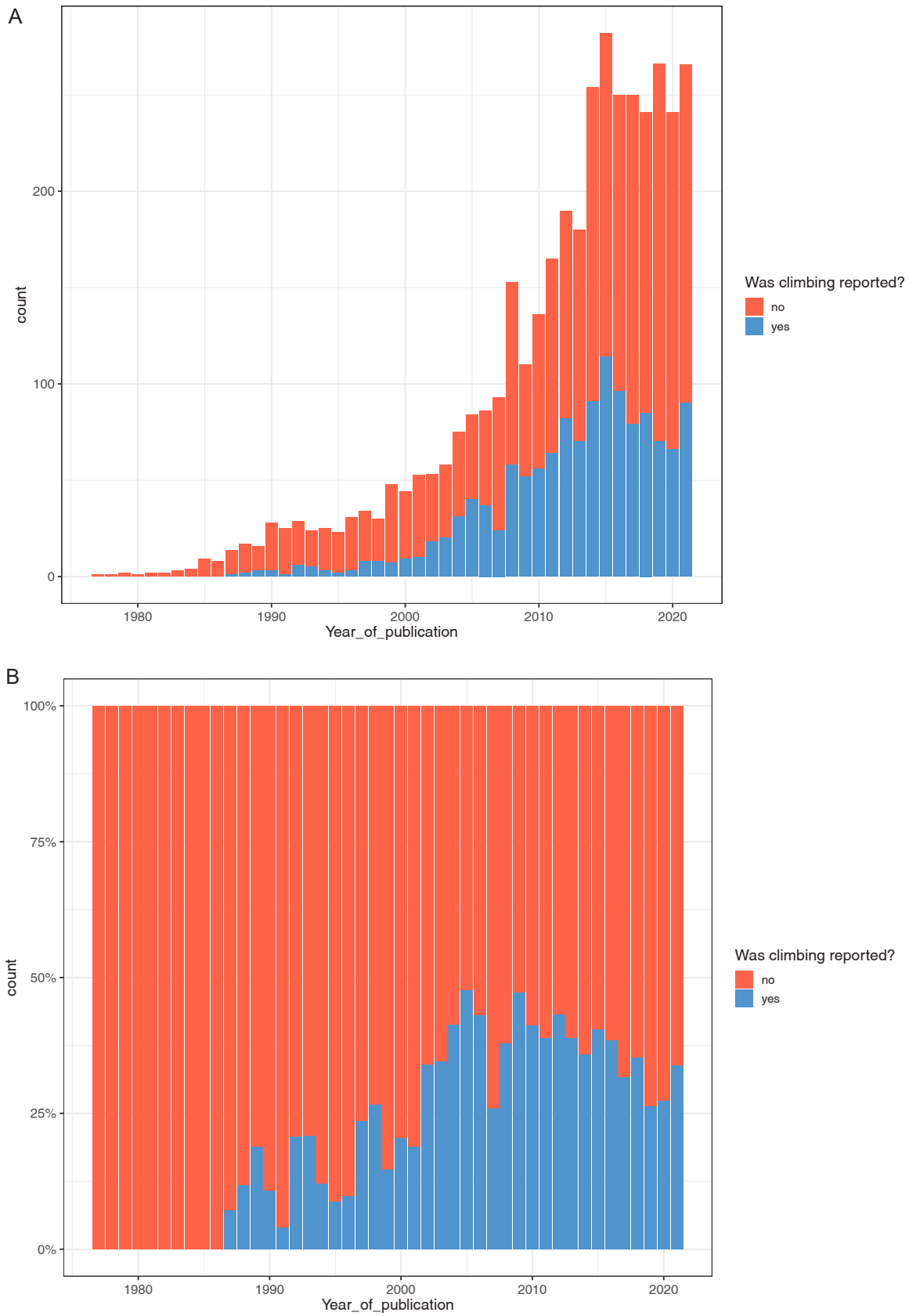

Fig. 10. Reporting of the outcome parameter "climbing" A) absolute numbers of publications B) percentages.

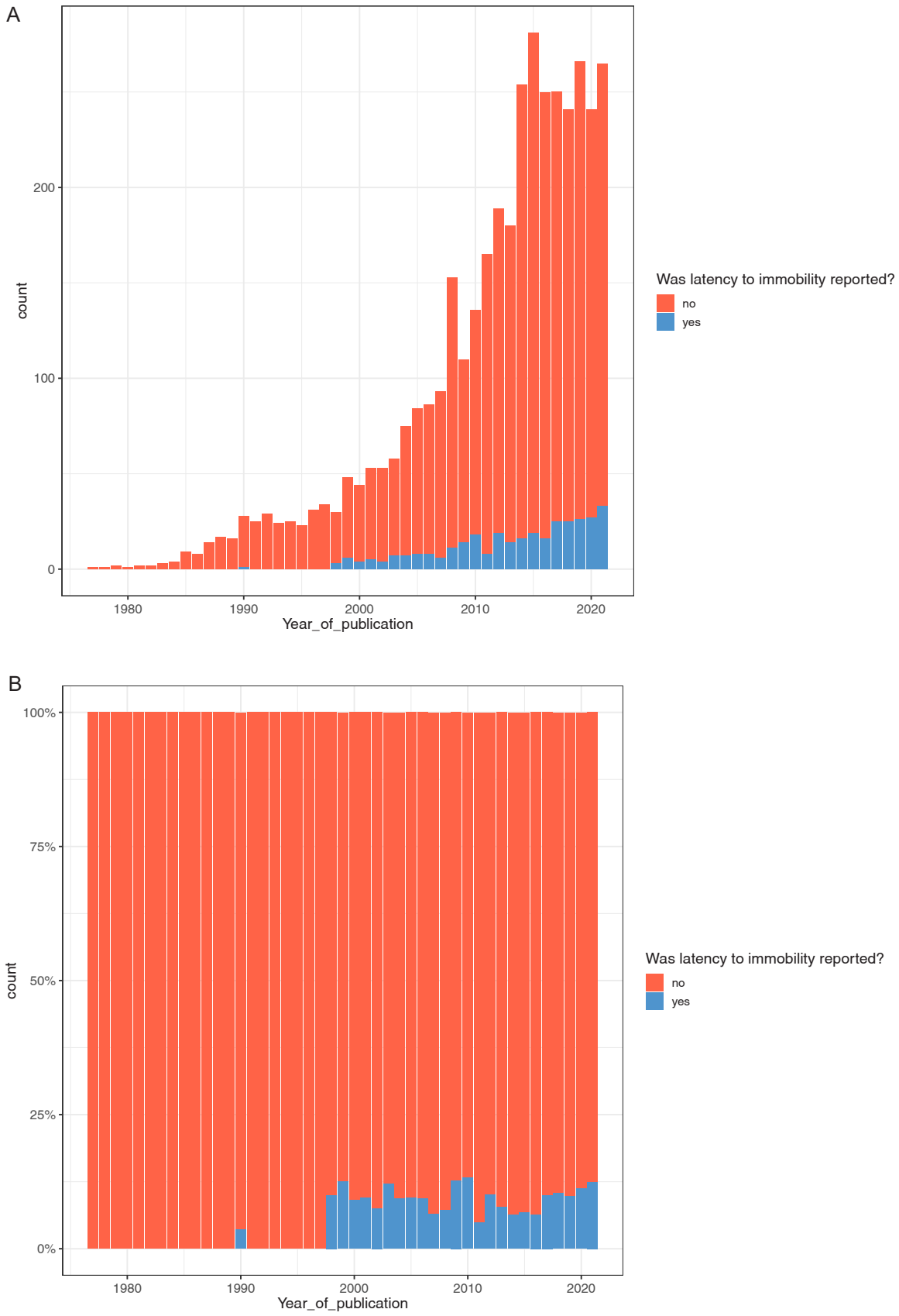

Fig. 11. Reporting of the outcome parameter "latency to first immobility" A) absolute numbers of publications B) percentages.

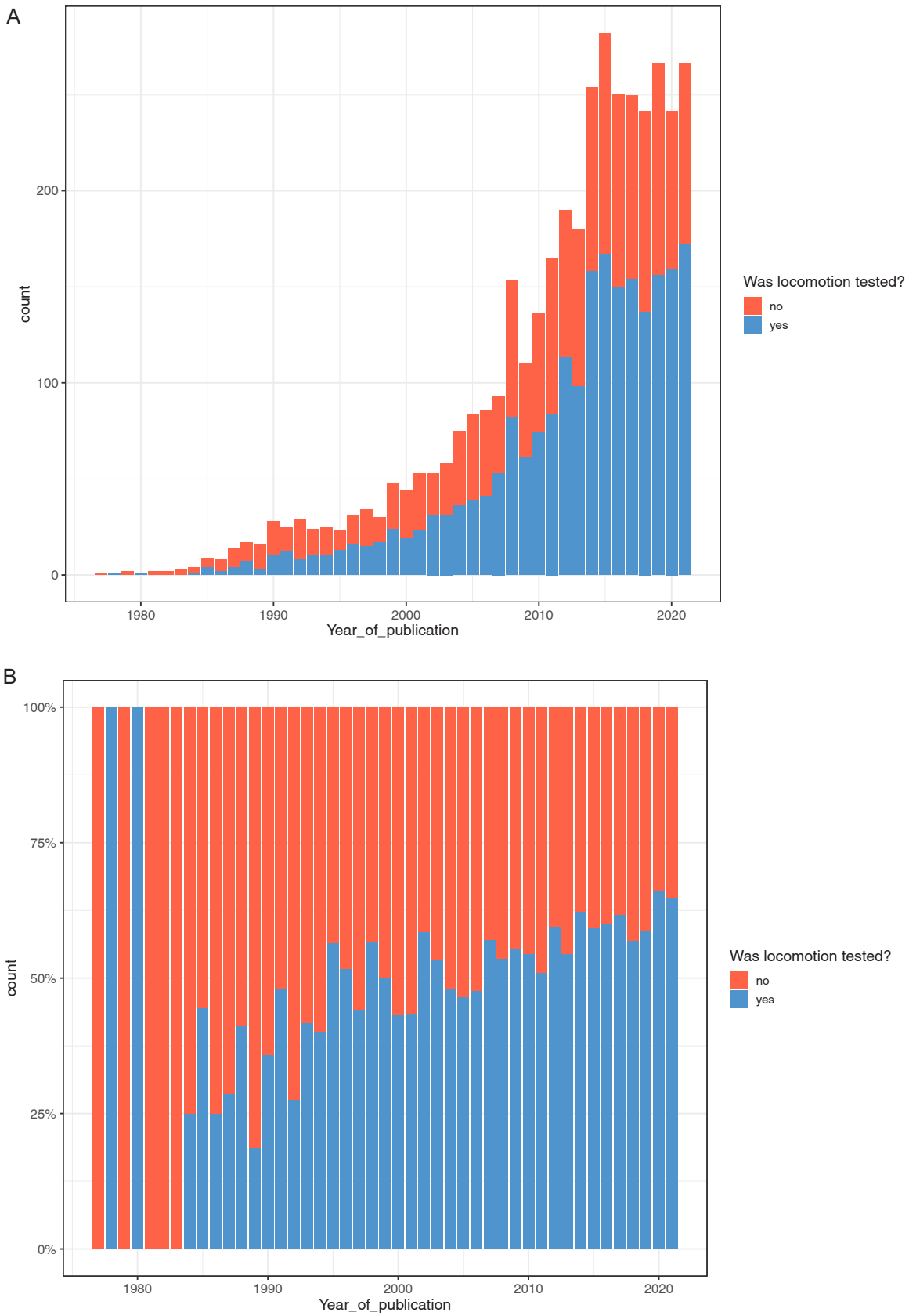

Fig. 12. Reporting of the outcome parameter "locomotion" A) absolute numbers of publications B) percentages.

publication, we will dive into which international guidelines were described as relevant in the publications. As a main purpose of the FST is to screen for antidepressant effects, we would have expected that many of the experiments would have received support or funding from industry, which could potentially bias the results. Our observation that the number of published experiments funded by industry stayed relatively low, while the total number of publications increased substantially, may suggest that industry studies are frequently not published. However, it also shows that the interest in this test in the scientific community is large, mirroring the increase in prevalence of depression in the world population.

The choice of strain appears to be very focused on Sprague Dawley and Wistar, both of which allow for meaningful experiments with two-sided statistical testing, as these two breeds display baseline behavior without ceiling or floor effects in terms of immobility [26–28].

The steady increase of the use of female animals in this test is a very positive, albeit slow, development, as it is only sensible to perform the screening in an animal population that includes the predominantly affected part of the human population to effectively increase the applicability of results [73,74]. Armario et al. [27] disarmed the argument of increased variance as they found no relevant differences between the sexes of several tested strains, even though the group sizes were fairly small (n = 5–7). Furthermore, it has been shown that when females are group-housed in the proestrous and estrous phase, and in the metestrous and diestrous phases of the cycle, their behavior in the FST was comparable [75–77]. D'Souza [78] had a closer look at the estrous cycle and found a phase-specific effect on baseline immobility in Wistars, but not in Wistar-Kyoto. But also in their set-up, the variances within a cycle phase were low. Therefore, to avoid a possible variance introduced through the different cycle phases, the advice would be the above-mentioned cycle synchronization. On the other hand, this again shows the importance of including all stages of the cycle within an experiment, to facilitate the bench-to-bedside process by including the effect of the estrous cycle on antidepressant efficacy as accurately as possible [77]. Further analyses of the k = 329 studies which included both female and male rats, to analyze variance of FST-outcomes between the sexes at the meta-level, are planned. If you are interested in sex differences in FST outcomes and/or their variance for your specific experiment, you can easily select the primary studies which analyzed both sexes from our spreadsheet [24] and choose the studies pertinent to your specific research question and experimental design for closer inspection.

Our observation of a slight decrease in group sizes was an unexpected finding due to the anticipated higher variance in the increasingly implemented female populations and the focus on experimental designs with a higher statistical power. On the other hand, possibly the emphasis on the 3Rs principle of reduction, refinement and replacement balanced out this expected increase.

The evolution of the currently approved antidepressive control substances as classified in Stahl's Essential Psychopharmacology [61], is also a very intriguing aspect. Imipramine, a tricyclic antidepressant (TCA) originally tested for possible antipsychotic effects, was a serendipitous discovery by the Swiss psychiatrist Robert Kuhn [79], introduced for medical use in 1957, and therefore it was already quite established when Porsolt published the FST in 1977 [13]. Similarly, Desipramine, which is also a TCA, was first authorized for the treatment of depression in 1963 [80]. Fluoxetine, the first selective serotonin reuptake inhibitor, was clinically introduced to the market in 1986 with the brand name "Prozac" [81]. As this class of antidepressants has quite a few advantages over the tricyclics, especially with respect to safety and tolerability, it is no surprise that Fluoxetine landed at the top of the list of applied control substances. Ketamine, an uncompetitive N-Methyl--D-Aspartate receptor antagonist, was also developed in the 1960s, as a derivative of phencyclidine for anesthesia with fewer hallucinogenic effects [82]. Its clinical antidepressant effect, however, was discovered much later [83], and it was only released to the market as an antidepressant nasal spray in 2019 [62]. Nonetheless, it already ranks within the top four control substances. The reason for that is that Ketamine is deemed a prototype for a new category of rapid-acting antidepressants, as it appears to have robust effects even in treatment resistant patients [62], although it is potentially also accompanied by a high level of dependency. This discovery has stimulated a recent wave of research into the neurobiological mechanisms underlying depressive disorders and their therapy, based on the molecular and cellular targets of ketamine [84].

The most commonly used water temperatures still largely adhere to the original protocol of Porsolt [12], where a temperature of 25 °C was used, even though it has been demonstrated that temperatures of 25 °C and below lead to hypothermia and increased, prolonged stress [85–87]. A water temperature of 30 °C has been shown to significantly reduce physical stress, while still reliably showing Desipramine induced anti-depressant activity [88]. However, in the k = 3721 FST studies published after 1993, when Abel et al. first published a comparative study including warmer water [87], only k = 201 (5.4%) used warmer water. Further meta-analyses of the sensitivity of the FST with higher temperatures, and an evaluation of the stress effect of lower water temperature in different experimental settings within our extracted dataset, are also planned.

The description of the age and weight of the animals in the included publications was rather poor, even though these parameters can significantly affect the FST outcomes [89]. It is therefore crucial for the aspired replicability of results to report these critical parameters, and we recommend emphasizing this complete reporting within the peer review process. While the one day protocol is commonly used in mice, it only started being implemented in rats after 1986 [90], at a stable low level of use over time. As mentioned in the introduction, some strains such as the Flinders Sensitive [25] and Wistar-Kyoto [26] do not necessarily benefit from having the preliminary exposure day. If it seems relevant to test over two days, though, Ecevitoglu et al. propose to calculate the ratio between the first 5 min of day one and day two, especially when assessing long-term treatment effects or inherent traits of the animals, as both days may be affected differently. This, however, only applies if there is no drug treatment between the first and the second swim [91].

A large majority of the included papers chose a normal light-dark cycle, testing the animals in their inactive phase. Theoretically it may be preferable to use a reversed dark-light cycle, where the animals would be tested during their active period. However, depending on the experienced side-effects, patients take their antidepressant medication in the morning and during the day. And Borsini et al. found that light may interfere with the dopaminergic receptors responsible for the anti-immobility effects induced by Desipramine [39], a discovery that was later supported by Pinter et al. [37]. Therefore, testing during the light phase seems to be the better approach for relevant translational results, unless a strain is tested that is prone to ceiling effects regarding immobility.

While Porsolt only reported the total immobility time as relevant outcome parameter in his original rat protocol [12], over time, several additional outcomes have received interest. For example, we observed a clear increase in the reporting of climbing and swimming as additionally observed parameters from 2002 onwards. We hypothesize that this is specific for testing SSRIs, which seem to have selective effects on these additional readouts [19]. Furthermore, it would be very interesting to see, if other uncommon outcome parameters, such as the latency of immobility (which was only observed in 330 studies) and diving (reported in 176 studies) also display different sensitivities to different interventions. As the effort is relatively small yet the added value may be high, assessing additional outcome measures within an experiment is advisable. As described earlier, our supplementary spreadsheet [24] can be used for easy retrieval of all relevant studies, also for those analyzing a specific outcome. By collating all relevant results pertinent to a specific experiment like this, the FST can be used to its maximum potential, in line with the 3Rs principle.

Locomotion is another factor that is increasingly implemented in test batteries comprising multiple behavioral tests to analyze depression-linked symptoms, not lastly because psychomotor retardation is also a marker for depression [92]. The increase of locomotor testing is an especially important evolution, because stimulants, sedatives, and motor-impairing compounds can additionally lead to false negatives and false positive results in the FST and other tests [93]. Optimally, multiple tests for antidepressant activity are performed in parallel, which either do not rely on locomotion or correct for it [94], as it has been shown that there is a correlation [95]. An example of a test not depending on locomotion was described by Williams et al. [96], who found that air-puffs induced 22-kHz ultrasonic vocalizations, which parallel immobility in the FST.

A limitation of this review is that it only analyzed the set-up used for rats, even though mice are at least equally important as rats for drug screening and for analyzing the effect of genetic mutations. However, including mice surpassed both our screening capacity and the desired level of complexity for this publication. The FST for mice is an imperative topic to address in a separate review. Another caveat is that all data were extracted by a single extractor, and that due to the size of the review, we could only perform a formal quality check of 5–10% of the included papers. Consequently, our data file may contain small errors. When we report extreme values, these were checked against the original publication, so these potential errors should not affect the conclusions of this paper. The inherent disadvantage of the comparatively superficial approach of Systematic Mapping Reviews [97] is surpassed by the grand advantage of the vast volume of literature we were able to include and synopsize with this approach. Further in-depth analyses, comprising meta-analyses of subsets of the screened literature, are planned by us within the scope of future publications. This review and the database included could also be the baseline and starting point for similar projects by interested readers [98].

In summary, for the time being, the FST is necessary and evidence-based, yet it needs to be employed with raised awareness of the details in the set-up and its limitations as a depression test due to its low construct validity. With this publication, which is the first to provide a full overview of the rat FST experimental set-ups and designs that have been published since its invention, we aim to support this process. Additionally, the searchable database allows researchers to easily retrieve the most relevant literature, facilitating a reduction of experiments or their refinement on the basis of evidence-based decision making.

Funding

This work was funded by the DFG FOR2591 [project GA-427/12-2 to PG, LE-2356/5-1 to LL, and BL-953/11-1 to AB].

CRediT authorship contribution statement

Original idea: PG, CB, ASM, CHCL, AB. Design of the review: CB, CHCL, ASM, PG. Systematic literature search: CB, CHCL. Extraction: CB, NP, LB, AMI, GAV, LL, RA, CA, FT, FM, BV. Analysis: CB, CHCL. Writing: CB, CHCL. Revision / editing: CB, CHCL, LB, PG, AMI, LL, ASM, AB. Funding / resources: PG, AB, LL.

Declaration of Competing Interest

None.

Acknowledgments

We thank Roberto Aldeghi, Corrado Antognini, Federica Terraneo, Federico Milan, and Barbara Vollmayr for their incredible support in the screening and extraction process, as well as Isabell Stein and her team for retrieving the vast number of articles for screening.

References

[1] WHO, Report of the Secretariat: Global Burden of Mental Disorders and the Need for A Comprehensive, Coordinated Response from Health and Social Sectors at the Country Level. Executive Board EB 130/9, 2011.
[2] GBD 2019 Mental Disorders Collaborators, Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet Psychiatry 9 (2) (2022) 137–150.
[3] C.K. Ettman, et al., Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic, JAMA Netw. Open 3 (9) (2020), e2019686.
[4] E. Malatynska, et al., Assessing activity onset time and efficacy for clinically effective antidepressant and antimanic drugs in animal models based on dominant-submissive relationships, Neurosci. Biobehav. Rev. 31 (6) (2007) 904–919.
[5] X. Gonda, et al., Novel antidepressant drugs: beyond monoamine targets, CNS Spectr. 28 (1) (2023) 6–15.
[6] M.H. Trivedi, et al., Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice, Am. J. Psychiatry 163 (1) (2006) 28–40.
[7] T.R. Insel, Beyond efficacy: the STAR*D trial, Am. J. Psychiatry 163 (1) (2006) 5–7.
[8] G.H. Vazquez, et al., Efficacy and tolerability of combination treatments for major depression: antidepressants plus second-generation antipsychotics vs. esketamine vs. lithium, J. Psychopharmacol. 35 (8) (2021) 890–900.
[9] S. Shorey, E.D. Ng, C.H.J. Wong, Global prevalence of depression and elevated depressive symptoms among adolescents: a systematic review and meta-analysis, Br. J. Clin. Psychol. 61 (2) (2022) 287–305.
[10] S.E. Murphy, et al., The knowns and unknowns of SSRI treatment in young people with depression and anxiety: efficacy, predictors, and mechanisms of action, Lancet Psychiatry 8 (9) (2021) 824–835.
[11] J.R. Homberg, et al., The continued need for animals to advance brain research, Neuron 109 (15) (2021) 2374–2379.
[12] R.D. Porsolt, M. Le Pichon, M. Jalfre, Depression: a new animal model sensitive to antidepressant treatments, Nature 266 (5604) (1977) 730–732.
[13] R.D. Porsolt, A. Bertin, M. Jalfre, Behavioral despair in mice: a primary screening test for antidepressants, Arch. Int. Pharmacodyn. Ther. 229 (2) (1977) 327–336.
[14] J.F. Cryan, A. Markou, I. Lucki, Assessing antidepressant activity in rodents: recent developments and future needs, Trends Pharmacol. Sci. 23 (5) (2002) 238–245.
[15] F. Borsini, A. Meli, Is the forced swimming test a suitable model for revealing antidepressant activity? Psychopharmacology 94 (2) (1988) 147–160.
[16] M. Bianchi, et al., Forced swimming test and fluoxetine treatment: in vivo evidence that peripheral 5-HT in rat platelet-rich plasma mirrors cerebral extracellular 5-HT levels, whilst 5-HT in isolated platelets mirrors neuronal 5-HT changes, Exp. Brain Res. 143 (2) (2002) 191–197.
[17] K.G. Commons, et al., The rodent forced swim test measures stress-coping strategy, not depression-like behavior, ACS Chem. Neurosci. 8 (5) (2017) 955–960.
[18] M.L. Molendijk, E.R. de Kloet, Coping with the forced swim stressor: current state-of-the-art, Behav. Brain Res. 364 (2019) 1–10.
[19] M.J. Detke, M. Rickels, I. Lucki, Active behaviors in the rat forced swimming test differentially produced by serotonergic and noradrenergic antidepressants, Psychopharmacology 121 (1) (1995) 66–72.
[20] O.V. Bogdanova, et al., Factors influencing behavior in the forced swim test, Physiol. Behav. 118 (2013) 227–239.
[21] V. Castagne, et al., Rodent models of depression: forced swim and tail suspension behavioral despair tests in rats and mice, Curr. Protoc. Pharmacol. (2010). Chapter 5: p. Unit 5 8.
[22] A. Gururajan, et al., The future of rodent models in depression research, Nat. Rev. Neurosci. 20 (11) (2019) 686–701.
[23] C.R. Hooijmans, M. Ritskes-Hoitinga, Progress in using systematic reviews of animal studies to improve translational research, PLoS Med. 10 (7) (2013), e1001482.
[24] C. Brandwein, C.H.C. Leenaars [dataset], A Systematic Mapping Review of the Evolution of the Rat Forced Swim Test: Protocols and Outcome Parameters, OSF (2023). osf.io/m3xdb, DOI 10.17605/OSF.IO/M3XDB.
C. Brandwein, C.H.C. Leenaars [dataset], A Systematic Mapping Review of the Evolution of the Rat Forced Swim Test: Protocols and Outcome Parameters, OSF (2023). osf.io/m3xdb, DOI 10.17605/OSF.IO/M3XDB.
[25] J. Gericke, et al., An acute dose-ranging evaluation of the antidepressant properties of Sceletium tortuosum (Zembrin(R)) versus escitalopram in the Flinders Sensitive Line rat, J. Ethnopharmacol. 284 (2022), 114550.
[26] P.A. Rittenhouse, et al., Amplified behavioral and endocrine responses to forced swim stress in the Wistar-Kyoto rat, Psychoneuroendocrinology 27 (3) (2002) 303–318.
[27] A. Armario, A. Gavalda, J. Marti, Comparison of the behavioural and endocrine response to forced swimming stress in five inbred strains of rats, Psychoneuroendocrinology 20 (8) (1995) 879–890.
[28] C. Lopez-Rubalcava, I. Lucki, Strain differences in the behavioral effects of antidepressant drugs in the rat forced swimming test, Neuropsychopharmacology 22 (2) (2000) 191–199.
[29] R.C. Kessler, Epidemiology of women and depression, J. Affect. Disord. 74 (1) (2003) 5–13.
[30] J.M. Gorman, Gender differences in depression and response to psychotropic medication, Gend. Med. 3 (2) (2006) 93–109.
[31] R.C. Turner, et al., Effects of aging on behavioral assessment performance: implications for clinically relevant models of neurological disease, J. Neurosurg. 117 (3) (2012) 629–637.

[32] E.L. Abel, Behavioral and physiological effects of different water depths in the forced swim test, Physiol. Behav. 56 (2) (1994) 411–414.
[33] J.M. De Pablo, et al., Learned immobility explains the behavior of rats in the forced swimming test, Physiol. Behav. 46 (2) (1989) 229–237.
[34] M.J. Detke, I. Lucki, Detection of serotonergic and noradrenergic antidepressants in the rat forced swimming test: the effects of water depth, Behav. Brain Res. 73 (1–2) (1996) 43–46.
[35] I. Lucki, The forced swimming test as a model for core and component behavioral effects of antidepressant drugs, Behav. Pharmacol. 8 (6–7) (1997) 523–532.
[36] J.F. Cryan, M.E. Page, I. Lucki, Differential behavioral effects of the antidepressants reboxetine, fluoxetine, and moclobemide in a modified forced swim test following chronic treatment, Psychopharmacology 182 (3) (2005) 335–344.
[37] O. Pinter, et al., Do stress hormones connect environmental effects with behavior in the forced swim test? Endocr. J. 58 (5) (2011) 395–407.
[38] K. Holl, et al., Heterogeneous stock rats: a model to study the genetics of despair-like behavior in adolescence, Genes Brain Behav. 17 (2) (2018) 139–148.
[39] F. Borsini, et al., Seasonal and circadian variations of behavioural response to antidepressants in the forced swimming test in rats, Behav. Pharmacol. 1 (5) (1990) 395–401.
[40] V. Castagne, Early behavioral screening for antidepressants and anxiolytics, Drug Dev. Res. 67 (2006) 729–742.
[41] F.S. Hall, et al., The effects of social isolation on the forced swim test in Fawn hooded and Wistar rats, J. Neurosci. Methods 79 (1) (1998) 47–51.
[42] S. Reardon, Depression researchers rethink popular mouse swim tests, Nature 571 (7766) (2019) 456–457.
[43] D.K. Yang, et al., Aralia continentalis kitagawa extract attenuates the fatigue induced by exhaustive exercise through inhibition of oxidative stress, Antioxidants 9 (5) (2020).
[44] F. Al-Hashem, et al., Exhaustive exercise and vitamins C and E modulate thyroid hormone levels at low and high altitudes, EXCLI J. 11 (2012) 487–494.
[45] L. Becker, et al., Evidence-Based Severity Assessment of the Forced Swim Test in the Rat (Submitted for publication).
[46] D. Moher, et al., Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009), e1000097.
[47] R.B.M. de Vries, et al., A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies, Evid. Based Preclin. Med. 2 (1) (2015), e00007.
[48] R. StudioTeam, RStudio: Integrated Development Environment for R, RStudio, PBC, Boston, MA, 2020.
[49] H. Wickham, J. Bryan, readxl: Read Excel Files, 2019.
[50] H. Wickham, et al., dplyr: A Grammar of Data Manipulation, 2023.
[51] D. Chaltiel, crosstable: Crosstables for Descriptive Analyses, R Package Version 0.5.0, 2022.
[52] H. Wickham, D. Vaughan, M. Girlich, tidyr: Tidy Messy Data, R Package Version 1.3.0, 2023.
[53] M. Dowle, A. Srinivasan, data.table: Extension of `data.frame`, R package version 1.14.8, 2023.
[54] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer Verlag, New York, 2016.
[55] A. South, rworldmap: A New R package for Mapping Global Data, R J., 2011.
[56] I. Feinerer, K. Hornik, Text Mining Infrastructure in R, J. Stat. Softw., 2008.
[57] I. Feinerer, K. Hornik, tm: Text Mining Package, R Package Version 0.7-8, 2020.
[58] C.R. Hooijmans, et al., SYRCLE's risk of bias tool for animal studies, BMC Med. Res. Methodol. 14 (2014) 43.
[59] S. van der Mierden, et al., Intracerebral microdialysis of adenosine and adenosine monophosphate - a systematic review and meta-regression analysis of baseline concentrations, J. Neurochem. 147 (1) (2018) 58–70.
[60] D.H. Overstreet, Selective breeding for increased cholinergic function: development of a new animal model of depression, Biol. Psychiatry 21 (1) (1986) 49–58.
[61] S. Stahl, Stahl's Essential Psychopharmacology: Neuroscientific Basis and Practical Applications, fifth ed., Cambridge University Press, Cambridge, 2021.
[62] K. Hashimoto, Rapid-acting antidepressant ketamine, its metabolites and other candidates: a historical overview and future perspective, Psychiatry Clin. Neurosci. 73 (10) (2019) 613–627.
[63] R.N. El-Naga, H.I. Ahmed, E.N. Abd Al Haleem, Effects of indole-3-carbinol on clonidine-induced neurotoxicity in rats: Impact on oxidative stress, inflammation, apoptosis and monoamine levels, Neurotoxicology 44 (2014) 48–57.
[64] H. Khatoon, R. Ikram, G. Abbas, Neuropharmacological effects of camel milk related to modulation of biogenic amines in experimental animals, Pak. J. Pharm. Sci. 32 (6) (2019) 2633–2641.
[65] Q. Zhang, et al., Effects of varicocele on DNA methylation pattern of H19 and Snrpn gene in spermatozoa and behavioural characteristics of adult rat offspring, Andrologia 49 (1) (2017).
[66] H.H. Gao, Biobehavioral effects produced by paternal sleep disturbances, Sleep. Biol. Rhythms 13 (2015) 235–241.
[67] J. Fernandez-Pardal, L.A. Hilakivi, Effect of low alcohol dose on behavioral "despair" in rats neonatally treated with antidepressant drugs, Alcohol 6 (2) (1989) 93–95.
[68] F.C. Colpaert, et al., Memory of an operant response and of depressed mood retained in activation states of 5-HT(1A) receptors: evidence from rodent models. Behav. Brain Res. 117 (1–2) (2000) 41–51.
[69] H. Nishimura, et al., Is immobility of rats in the forced swim test "behavioral despair, Physiol. Behav. 42 (1) (1988) 93–95.
[70] A. Nakasato, et al., Swim stress exaggerates the hyperactive mesocortical dopamine system in a rodent model of autism, Brain Res. 1193 (2008) 128–135.
[71] E.R. Trunnell, C. Carvalho, The forced swim test has poor accuracy for identifying novel antidepressants, Drug Discov. Today 26 (12) (2021) 2898–2904.
[72] M.L. Molendijk, E.R. de Kloet, Immobility in the forced swim test is adaptive and does not reflect depression, Psychoneuroendocrinology 62 (2015) 389–391.
[73] J.A. Clayton, F.S. Collins, Policy: NIH to balance sex in cell and animal studies, Nature 509 (7500) (2014) 282–283.
[74] S. Heidari, et al., Sex and gender equity in research: rationale for the SAGER guidelines and recommended use, Res. Integr. Peer Rev. 1 (2016) 2.
[75] P.J. Allen, et al., Chronic creatine supplementation alters depression-like behavior in rodents in a sex-dependent manner, Neuropsychopharmacology 35 (2) (2010) 534–546.
[76] N. Kokras, et al., Behavioral sexual dimorphism in models of anxiety and depression due to changes in HPA axis activity, Neuropharmacology 62 (1) (2012) 436–445.
[77] N. Kokras, et al., Forced swim test: what about females? Neuropharmacology 99 (2015) 408–421.
[78] D. D'Souza, M. Sadananda, Estrous cycle phase-dependent changes in anxiety- and depression-like profiles in the late adolescent wistar-kyoto rat, Ann. Neurosci. 24 (3) (2017) 136–145.
[79] R. Kuhn, Treatment of depressive states with an iminodibenzyl derivative (G 22355), Schweiz. Med. Wochenschr. 87 (35–36) (1957) 1135–1140.
[80] J. Andersen, et al., Recent advances in the understanding of the interaction of antidepressant drugs with serotonin and norepinephrine transporters, Chem. Commun. (25) (2009) 3677–3692.
[81] K.R. Connolly, M.E. Thase, Emerging drugs for major depressive disorder, Expert Opin. Emerg. Drugs 17 (1) (2012) 105–126.
[82] B. Le Dare, et al., History of ketamine: an ancient molecule that is still popular today, Ann. Pharm. Fr. 80 (1) (2022) 1–8.
[83] R.M. Berman, et al., Antidepressant effects of ketamine in depressed patients, Biol. Psychiatry 47 (4) (2000) 351–354.
[84] J.H. Krystal, et al., Ketamine: a paradigm shift for depression research and treatment, Neuron 101 (5) (2019) 774–778.
[85] J.F. Taltavull, et al., Severe brain hypothermia as a factor underlying behavioral immobility during cold-water forced swim, Brain Res. 975 (1–2) (2003) 244–247.
[86] R.C. Drugan, et al., Impact of water temperature and stressor controllability on swim stress-induced changes in body temperature, serum corticosterone, and immobility in rats, Pharmacol. Biochem. Behav. 82 (2) (2005) 397–403.
[87] E.L. Abel, Physiological correlates of the forced swim test in rats, Physiol. Behav. 54 (2) (1993) 309–317.
[88] A.A. Walf, M.E. Rhodes, C.A. Frye, Antidepressant effects of ERbeta-selective estrogen receptor modulators in the forced swim test. Pharmacol. Biochem. Behav. 78 (3) (2004) 523–529.
[89] L. Martinez-Mota, A. Fernandez-Guasti, Testosterone-dependent antidepressant-like effect of noradrenergic but not of serotonergic drugs, Pharmacol. Biochem. Behav. 78 (4) (2004) 711–718.
[90] D.H. Overstreet, et al., Stress-induced immobility in rats with cholinergic supersensitivity, Biol. Psychiatry 21 (7) (1986) 657–664.
[91] A. Ecevitoglu, et al., Paw preference is associated with behavioural despair and spatial reference memory in male rats, Behav. Process. 180 (2020), 104254.
[92] G. Unal, R. Canbeyli, Psychomotor retardation in depression: a critical measure of the forced swim test, Behav. Brain Res. 372 (2019), 112047.
[93] S.C. Stanford, The Open Field Test: reinventing the wheel, J. Psychopharmacol. 21 (2) (2007) 134–135.
[94] D.A. Slattery, I.D. Neumann, J.F. Cryan, Transient inactivation of the infralimbic cortex induces antidepressant-like effects in the rat, J. Psychopharmacol. 25 (10) (2011) 1295–1303.
[95] C. Lino-de-Oliveira, T.C. De Lima, A. de Padua Carobrez, Structure of the rat behaviour in the forced swimming test, Behav. Brain Res. 158 (2) (2005) 243–250.
[96] A.M. Williams, et al., The effect of intermittent alcohol vapor or pulsatile heroin on somatic and negative affective indices during spontaneous withdrawal in Wistar rats, Psychopharmacology 223 (1) (2012) 75–88.
[97] C. Leenaars, et al., Reviewing the animal literature: how to describe and choose between different types of literature reviews, Lab. Anim. 55 (2) (2021) 129–141.
[98] A. Culina, et al., How to do meta-analysis of open datasets, Nat. Ecol. Evol. 2 (7) (2018) 1053–1056.