Quality, topics, and demographic trends of animal systematic reviews - an umbrella review

Bernard Friedrich Hild[1†], David Brüschweiler[1†], Sophia Theodora Katharina Hild[1], Julia Bugajska[1], Viktor von Wyl[2,3], Marianna Rosso[1], Kimberley Elaine Wever[4], Eva Furrer[1] and Benjamin Victor Ineichen[1,2,5*]

Abstract

Background Animal systematic reviews are critical to inform translational research. Despite their growing popularity, there is a notable lack of information on their quality, scope, and geographical distribution over time. Addressing this gap is important to maintain their effectiveness in fostering medical advancements.

Objective This study aimed to assess the quality and demographic trends of animal systematic reviews in neuroscience, including changes over time.

Methods We performed an umbrella review of animal systematic reviews, searching Medline and Embase for reviews until January 27, 2023. A data mining method was developed and validated to automatically evaluate the quality of these reviews.

Results From 18'065 records identified, we included 1'358 animal systematic reviews in our study. These reviews commonly focus on translational research but with notable topical gaps such as schizophrenia, other psychiatric disorders, and brain tumours. They originate from 64 countries, with the United States, China, the UK, Brazil, and Iran being the most prolific. The automated quality assessment indicated high reliability, with F1-scores over 80% for most criteria. Overall, the reviews were of high quality and the quality improved over time. However, many systematic reviews did not report a pre-registered study protocol. Reviews with a pre-registered protocol generally scored higher in quality. No significant differences in quality were observed between countries.

Conclusion Animal systematic reviews in neuroscience are of overall of high quality. Our study highlights specific areas for enhancement such as the recommended pre-publication of study protocols. It also identifies under-represented topics that could benefit from further investigation to inform translational research. Such measures can contribute to the effective translation of animal research findings to clinical applications.

Keywords Translational research, Systematic review, Animal research, Neuroscience, Animal welfare, Evidence map, Automation, Risk of bias

†Bernard Friedrich Hild and David Brüschweiler contributed equally and share the first authorship.

*Correspondence:
Benjamin Victor Ineichen
Benjamin.ineichen@uzh.ch

Full list of author information is available at the end of the article

Introduction

A systematic review is a comprehensive synthesis of existing research on a specific topic, typically including a comprehensive search, explicit in- and exclusion criteria, critical appraisal, and structured analysis of relevant studies to summarize the findings and draw evidence-based conclusions [1–3]. Originally developed for clinical research, systematic reviews are now increasingly used in preclinical fields [4, 5]. Their increasing use in this area can enhance our understanding of disease mechanisms and foster the transition from laboratory research to clinical application. This can provide insights without necessitating new animal experiments [6–9]. For example, a systematic review found that treatments tested in clinically relevant animal models (e.g., stroke drugs tested in older animals with comorbidities) are more likely to succeed in human trials, while those tested in less realistic models often fail [10]. Similarly, a pancreatitis clinical trial using probiotics showed no benefit and even higher mortality [11], likely because a systematic assessment of earlier animal studies showed different probiotics and administration regimens, highlighting discrepancies between animal models and clinical trials [12]. Systematic reviews also improve research transparency, show missing evidence, and identify needs and designs for future studies.

However, the utility of animal systematic reviews to inform human health depends on their rigour and relevance. Recent findings indicate a prevalence of overall low-quality clinical [13] but also preclinical systematic reviews [14–17]. Yet there is a lack of detailed assessment of preclinical systematic review rigour, which is increasingly challenging due to the rapid growth of biomedical literature [18]. Additionally, understanding country-specific differences in systematic review quality could guide targeted training efforts aimed at improving review standards, as advocated by international collaborations like SYRLCE and CAMARADES [19, 20]. This would not only support researcher training but also reveal trends in the most prolific countries for systematic reviews. Finally, it remains unclear to what extent preclinical systematic reviews align with topics relevant to human health, particularly prevalent diseases. Strengthening this alignment is essential for translational science to advance clinical care [21]. Further, mapping these topics could identify underexplored areas within preclinical evidence synthesis.

To address these issues, we have conducted an umbrella systematic review aimed at mapping animal systematic reviews in neuroscience, a major field within biomedical research. Our analysis aimed to answer three questions: (1) What research topics have been addressed by animal systematic reviews in neuroscience, including their translational focus? (2) Which countries are the leading producers of animal systematic reviews, and how has this changed over time? (3) What is the overall quality of these systematic reviews in animal neuroscience, does it vary across different countries, and is there an improvement in quality over time?

Materials and methods

Study registration

We registered a study protocol for our umbrella review [22] on the Open Science Framework platform (OSF, https://osf.io/wx5ta/) on November 25 2023. This review used the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines for reporting [23].

Information sources and search strategy

To identify all available animal systematic reviews in neuroscience, we searched for studies published from database inception up to January 27, 2023, on Embase and Medline (both via Ovid). The search string was created in Medline and translated to Embase (Supplementary Data). In brief, the search string was comprised of a component for systematic reviews and meta-analysis and a component for neuroscience, and was limited to animal studies by employing the SYRCLE animal filter [24].

Eligibility criteria

Inclusion criteria

In line with the JBI Handbook for umbrella reviews [25], we included both systematic reviews and meta-analyses summarizing animal studies in neuroscience, as meta-analyses often accompany systematic reviews in this field [3]. These studies required at least two of the following three items to be fulfilled: (1) an explicit mention of the terms "systematic review" or "meta-analysis" in the title or abstract, or by (2) mentioning a systematic literature search in at least two databases in the abstract, and by (3) mentioning adherence to systematic review guidelines in the abstract. We selected these criteria deliberately, as authors are likely to use the terms "systematic review" or "meta-analysis" for clarity and to follow established guidelines (e.g., Cochrane, CAMARADES, SYRCLE). Additionally, a systematic review's defining characteristic is its comprehensive literature search [3]. We also included systematic reviews synthesizing both animal and human primary data.

Exclusion criteria

Primary studies. Non-systematic reviews. Systematic reviews outside the neurosciences. Systematic reviews only including human, in vitro or in silico data.

Study selection process and (automated) data extraction

Two reviewers (DB and BFH) screened titles and abstracts of studies for their relevance in the web-based application Rayyan [26]. In accordance with the study protocol, we did not review the full texts of these studies due to the anticipated high volume of eligible studies. Subsequently, we extracted the following (meta) data: title, authors, publication year, journal, and number of authors. Data on study country and author keywords were retrieved by matching DOIs of respective publications with the Embase export. Additionally, based on the abstract, systematic reviews were manually categorized based on whether they focused on translational or purely basic research questions including whether systematic reviews included only animal studies or both animal and human studies.

We automatically extracted items related to the quality of systematic reviews from the full text. This approach was inspired by earlier suggestions [9] and further extended by additional criteria. These elements pertain either to how the systematic reviews were reported or to the quality of their methodological approach. The elements focusing on reporting include: (1) Was a study protocol drafted? (2) Was screening and/or extraction conducted by two or more reviewers? (3) Was a research question and/or study goal defined? (4) Were two or more literature databases searched? Items related to the methodological quality were: (5) Was a flowchart for study selection provided? (6) Was a conflict-of-interest statement provided? (7) Were in- and exclusion criteria reported? (8) Was a literature search date provided? (9) Was a literature search string provided? (10) Was a critical appraisal of included studies conducted? 11) Did the study mention any relevant systematic review guidelines, e.g., SYRCLE, CAMARADES, or PRISMA? For the sake of clarity, we will refer to all these items as quality items forward.

These items were automatically extracted from full texts of eligible studies using a custom-built R tool. This tool uses regular expressions, i.e., patterns of characters that define specific text matches, to match relevant keywords in the methods and results sections of the respective studies. For each of these items, we created libraries of regular expressions based on another umbrella review of 120 systematic reviews in translational biomedicine [27]. The tool segments each paper into sections (like results or methods), removes the 'references' section, and finally searches for matching regular expression patterns. To evaluate performance of this tool in this study, we manually examined a random 10% sample of the systematic reviews and calculated inter-rater agreement. Discrepancies were resolved by discussion. Both our regular expression libraries and the R tool are available at: https://osf.io/wx5ta/.

Data synthesis and analysis

For the analysis of covered topics, study keywords as provided by the authors were organized in descending order of frequency and aligned with a pre-established list of neurological/psychiatric conditions. In addition, each systematic review was manually classified by two independent reviewers into a specific disease according to the disease category from the Global Burden of Disease study for neurological [28] and mental disorders [29] and assigned to one or more of the following topics: therapeutic intervention, pathophysiology and mechanisms, diagnostic tools and biomarkers, and other. To assess how well a disease is covered by preclinical systematic reviews, we calculated the ratio of systematic review counts per disease to the respective disease prevalence (multiplied by 10,000), referred to as the systematic review-disease prevalence ratio. For the analysis of quality over time and growth rate, we focused our analysis on years with >10 systematic reviews being published per year. For the quality analysis, we assigned 1 point to each of the 11 items described above, allowing for a maximum score of 11 points and normalized the lowest and highest score to 0 and 1, respectively.

We summarized findings in narrative fashion and present descriptive statistics for demographic parameters. We conducted two statistical tests to compare quality scores, i.e., an ANOVA to compare the average systematic review quality per country and an unpaired t-test to compare the quality scores between systematic reviews with or without a pre-registered study protocol. RStudio (Version 2023.03.0, Build 386) running R Version 4.2.3 (2023-03-15 ucrt, "Shortstop Beagle") was used for all analyses and visualizations.

Results

Study selection and general study characteristics

We retrieved 18'065 unique records from our database search. After title and abstract screening, we included 1'358 studies (Fig. 1), with approximately half of them containing a meta-analysis (50%). Most systematic reviews had 4 to 5 authors (median 5, range 1–64).

Growth of the systematic review library

The first systematic review was published in 1997. Since 2007, the number of new systematic reviews in a year has grown increasingly, surmounting the number of systematic reviews published yearly from 5 in 2007 to 305 in 2022. The median global growth of systematic reviews per year was 26%.

Between 2018 and 2023, the systematic review library has more than tripled, surging from 417 systematic reviews at the start of 2018 to 1'331 SRs in 2022. Given the mean annual growth rate of 29% and assuming a steady growth, a doubling of the number of systematic
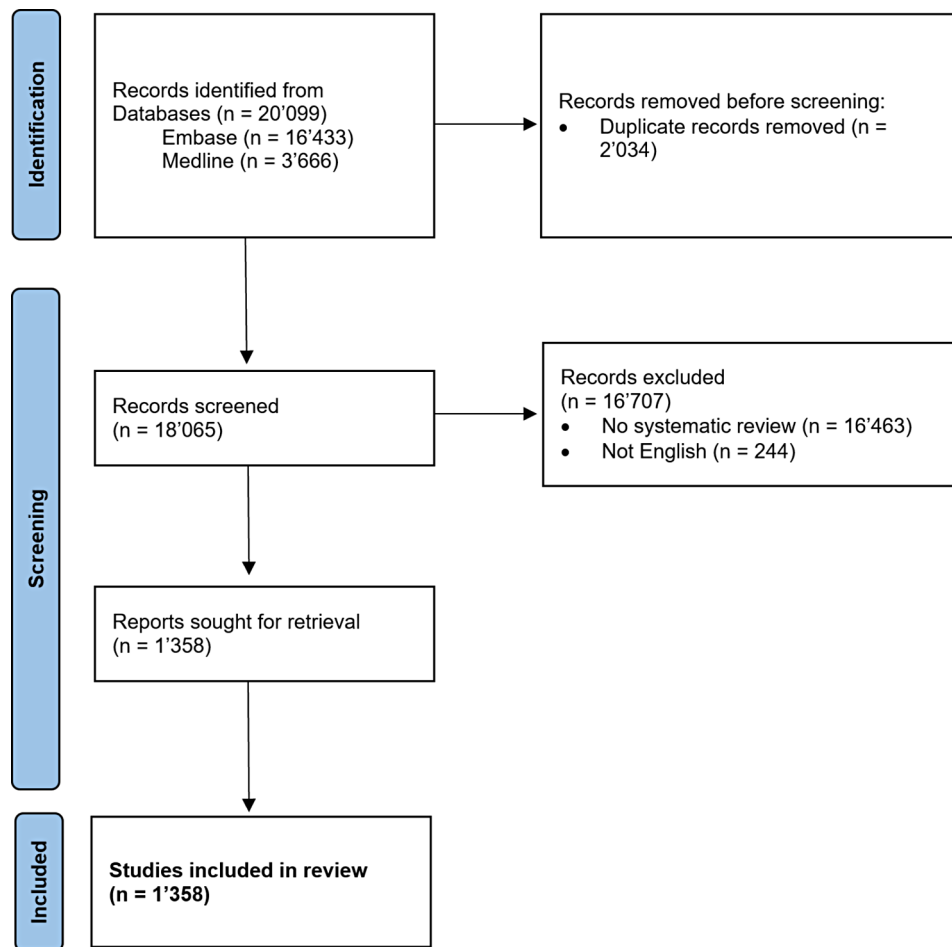
Fig. 1 Flow chart for study inclusion

reviews can be anticipated every 2.7 years (Supplementary Table 1).

Systematic review topics

53% of the systematic reviews included in our study addressed translational research questions (*n*=724). In contrast, relatively few systematic reviews focused on basic research (*n*=124, 9%). Along these lines, 938 systematic reviews included only animal studies (69%) and 420 systematic reviews included both animal and human studies (31%).

The most common research areas covered were neurodegenerative diseases (*n*=333, 25%), pharmacology (*n*=318, 23%), ischemia/cerebrovascular disease (*n*=310, 23%), and psychiatry (*n*=220, 16%), constituting over two thirds of the included systematic reviews (Fig. 2). Less frequently covered areas were neurobiology and physiology (*n*=124, 9%), neonatal and neurodevelopmental disorders (*n*=84, 6%), neuromuscular disorders (*n*=82, 6%), and neuro-gastroenterology (*n*=73, 5%).

The most frequently covered diseases were stroke (*n*=246, 18%), spinal cord injury (*n*=119, 9%),

Alzheimer's disease (*n*=116, 9%), traumatic brain injury (*n*=61, 4%), pain and Parkinson's disease (*n*=57 each, 4%), nervous system cancer (*n*=50, 4%), motor neuron disease (*n*=41, 3%), epilepsy (*n*=40, 3%), multiple sclerosis (*n*=38, 3%), and depression (*n*=33, 2%) (Fig. 3). The majority of systematic reviews focused on therapeutic interventions (65%), followed by studies on pathophysiological mechanisms (30%). Reviews with a primary focus on diagnostic approaches or biomarkers were less common (6%).

Finally, when correlating the number of systematic reviews to the prevalence of neurological and psychiatric diseases using global burden of disease data, certain topics appeared relatively overrepresented. These included motor neuron diseases (e.g., amyotrophic lateral sclerosis), which had the highest systematic review-disease prevalence ratio, as well as nervous system cancer, multiple sclerosis, spinal cord injury, and Guillain-Barré syndrome (Fig. 3). Conversely, many common psychiatric diseases such as depression, bipolar disorders, autism spectrum disorder, eating disorders, attention deficit hyperactivity disorder, anxiety disorders, and addiction,
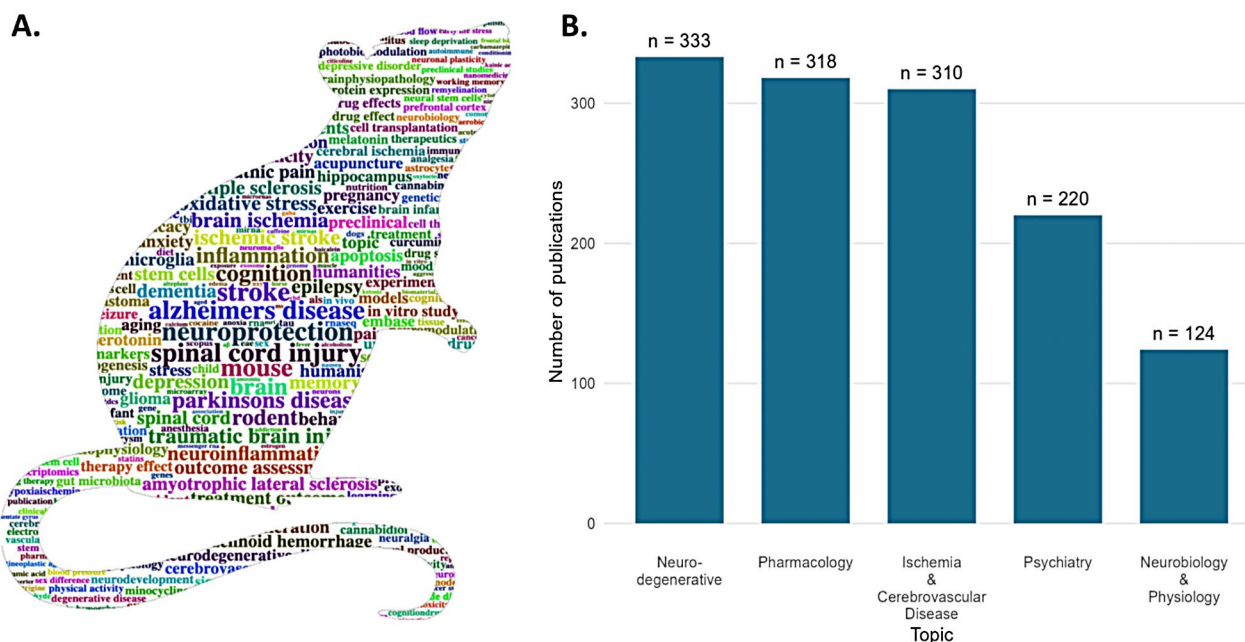
Fig. 2 Topics addressed by animal systematic reviews. Word cloud of covered topics (**A**) and topic quantification per research domain (**B**)

as well as conditions like diabetic neuropathy, pain, and migraine, were underrepresented. Notably, some prevalent disorders, including obsessive-compulsive disorder, Down syndrome, meningitis, and tension-type headache, had no identified preclinical systematic reviews at all.

Geography of systematic reviews

Included systematic reviews stemmed from 64 countries across all continents (Fig. 4).

Europe emerged as the most prolific producer of systematic reviews (n=507, 37% of total). The top three countries within Europe account for 54% of this output, with the UK (n=144), the Netherlands (n=62), and Italy (n=61) leading the way. Asia was the second most prolific continent (n=361, 26%). Notably, half of Asia's publications originate from China (n=182), with Iran also being a prolific country (n=87). The third most prolific continent was North America (n=301, 22%), with reviews mostly published be the United States (n=216, 72% of North American publications) and Canada (n=78, 26%). There were also systematic reviews from South America (n=101, 8%, predominantly represented by Brazil, n=89), Australia & Oceania (n=69 studies, 5%), and Africa (n=13, 1%, mostly from Egypt, Nigeria, and South Africa). Median annual growth rates for Europe, Asia, North America, South America, and Australia and Oceania were +20%, +44%, +21%, +38%, and +16%, respectively.

The most prolific countries, collectively covering nearly 50% of all SR, were the USA (n=216, 16% of global publications), China (n=182, 13%), the UK (n=144, 11%),

Brazil (n=89, 7%), Iran (n=87, 6%), and Canada (n=78, 6%) (Fig. 5A). Median annual growth rates for these countries were +22%, +42%, +14%, +33%, +70%, and +19%, respectively (Fig. 5B). This pattern remained consistent when analysing only systematic reviews without meta-analyses (Supplementary Figs. 1 and 2).

Quality of systematic reviews globally and over time

The automated quality mining function performed well in a random sample of the included references, with F1-scores well above 0.8 for most items (expect for whether the existence of a preregistered study protocol was reported in the systematic review, F1-score=0.72) (Supplementary Table 2).

The overall median quality score of systematic reviews (with "1" and "0" respectively signifying that all or none of the eleven quality items were fulfilled) was 0.818 (range 0.182–1) (Fig. 6A). The three most commonly fulfilled items, all related to reporting quality, were a clear statement of a research question (Fulfilled by 100% of included systematic reviews), the reporting of in- and exclusion criteria (100%), and reporting of a search date (93%). The three least checked items, all related to methodological quality, were the conduction of a critical appraisal of included primary studies (69%), the screening/extraction being conducted by two or more reviewers (59%), and the reporting of a study protocol (18%) (Fig. 6B).

Over time, the overall quality of systematic reviews significantly improved, with a median score of 0.51 (range 0.47–0.67) in the publication date range 2010–2012 and

	Ratio SR/prevalence	Prevalence	DALY	YLD	YLL	Deaths	SR total	SR therapeutic	SR mechanism	SR diagnosis	SR other
Motor neuron disease	1501.8	273	1040	57.9	983	39.1	41	17	21	2	9
Nervous system cancer	485.4	1030	9200	132	9070	264	50	36	11	5	9
Multiple sclerosis	201.1	1890	973	484	490	16.3	38	29	11	1	6
Spinal cord injury	77.3	15400	NA	4570	NA	NA	119	93	20	3	17
Guillain-Barre syndrome	60.2	166	NA	49	NA	NA	1	0	0	0	0
Fetal alcohol syndrome	58	517	NA	23.6	NA	NA	3	1	2	0	2
Parkinson's Disease	48.3	11800	7470	1670	5800	388	57	36	18	1	7
Neural tube defects	35.4	1130	5300	333	4970	56.4	4	4	0	0	0
Stroke	26.2	93800	160000	15200	145000	7250	246	176	45	17	32
Alzheimer Disease and other dementias	20.4	56900	36300	11600	24700	1950	116	79	45	7	16
Epilepsy	16.4	24400	14400	7760	6610	140	40	27	19	4	8
Traumatic brain injury	16.1	38000	NA	5490	NA	NA	61	42	29	5	12
Encephalitis	6.5	4640	4950	497	4460	92	3	2	0	0	1
Congenital birth defect	5	13900	NA	1450	NA	NA	7	3	3	1	2
Schizophrenia/psychosis	4.9	14200	NA	NA	NA	NA	7	6	6	0	8
Neonatal encephalopathy	4.8	18600	58600	4280	54300	604	9	7	3	0	0
Depression	1.9	170800	NA	NA	NA	NA	33	19	15	2	3
Bipolar disorder	1.6	24800	NA	NA	NA	NA	4	3	0	0	1
Autism spectrum disorder	1.5	61800	NA	11500	NA	NA	9	5	7	0	2
Eating disorder	1.2	8500	NA	NA	NA	NA	1	1	1	0	0
Attention deficit hyperactivity disorder	0.7	84800	NA	1030	NA	NA	6	2	2	0	4
Anxiety disorder	0.6	194900	NA	NA	NA	NA	12	3	7	1	1
Diabetic neuropathy	0.4	206000	NA	26300	NA	NA	9	7	3	0	0
Pain	0.2	2788000	NA	NA	NA	NA	57	34	21	4	9
Addiction	0.2	696000	NA	NA	NA	NA	12	8	4	1	2
Migraine	0.1	1160000	NA	43400	NA	NA	12	6	6	0	2
Obsessive compulsive disorder	0	700000	NA	NA	NA	NA	1	1	0	0	0
Cerebral malaria	0	996	NA	407	NA	NA	0	0	0	0	0
Cystic echinococcsis	0	15.1	NA	4.6	NA	NA	0	0	0	0	0
Down syndrome	0	1490	NA	135	NA	NA	0	0	0	0	0
Klinefelter syndrome	0	230	NA	3	NA	NA	0	0	0	0	0
Meningitis	0	7270	14500	603	13900	214	0	0	0	0	0
Neurocysticercosis	0	4360	1240	1150	81.3	1.6	0	0	0	0	0
Neurosyphilis	0	696	NA	64.9	NA	NA	0	0	0	0	0
Rabies	0	0.4	571	0.1	572	10.1	0	0	0	0	0
Tension-type headache	0	2010000	NA	4600	NA	NA	0	0	0	0	0
Tetanus	0	37.7	1340	3	1340	21.3	0	0	0	0	0
Other neurological diseases	NA	NA	4360	2010	2330	71.5	238	130	64	21	57

Fig. 3 Disease burden metrics and coverage of preclinical systematic reviews. This table presents disease burden metrics, including global disease prevalence (in thousands), disability-adjusted life-years (DALY), years lived with disability (YLD), years of life lost (YLL), and deaths. The metrics are based on disease categories from the Global Burden of Disease study for neurological [28] and mental disorders [29]. Additionally, the table includes the number of systematic reviews per disease, further categorized by focal topics: therapeutic interventions, pathophysiology and mechanisms, diagnostic tools, and biomarkers, and other. The first column displays the ratio of systematic reviews to disease prevalence (multiplied by 10,000) to indicate the relative over- or under-representation of systematic reviews for each disease. *Abbreviations: DALYs = disability-adjusted life-years; SR, systematic review; YLDs = years lived with disability; YLLs = years of life lost*
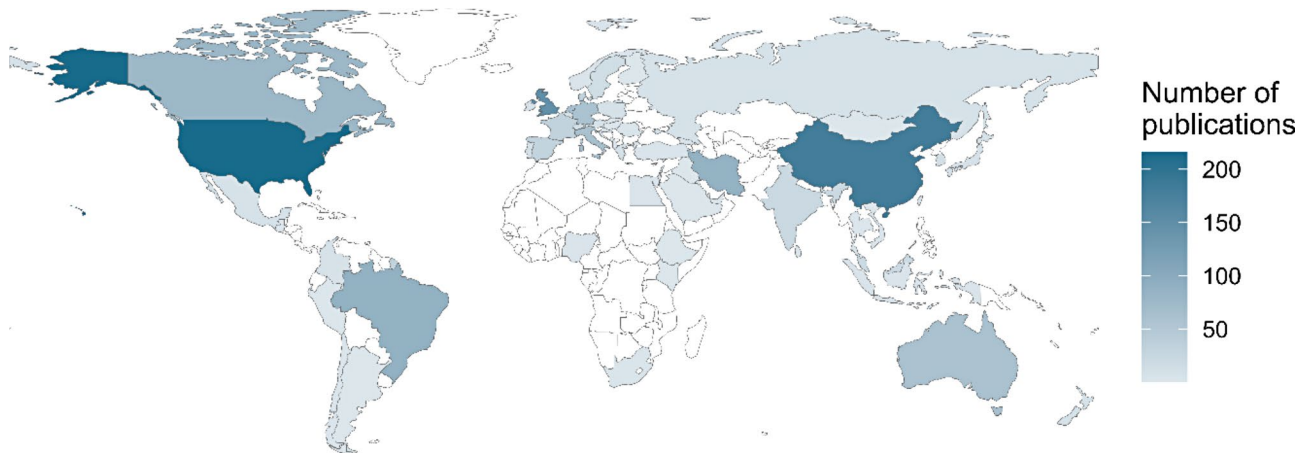

Fig. 4 World heat map of countries publishing animal systematic reviews. Animal systematic reviews stemmed from 64 countries across all continents
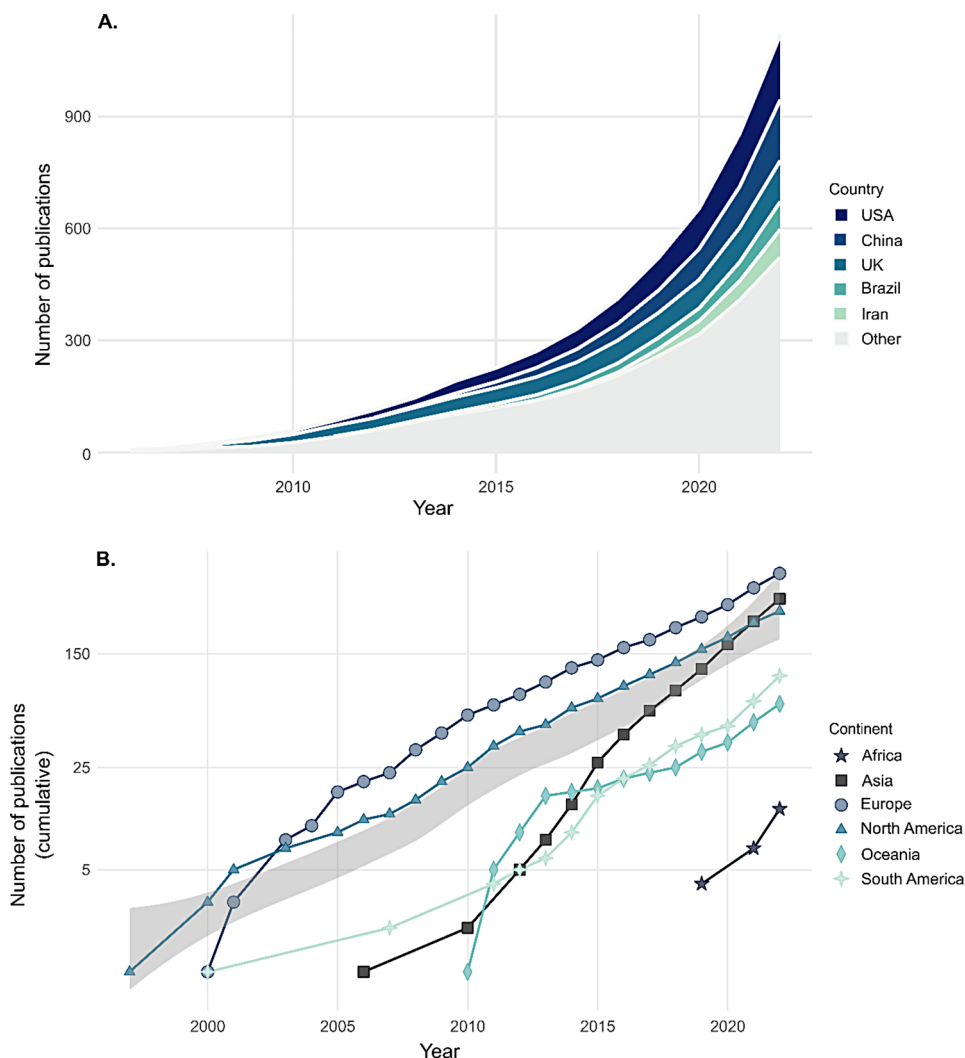
Fig. 5 Prolific countries publishing animal systematic reviews. The 5 most prolific countries were the USA, China, UK, Brazil, and Iran (**A**). Europe was the most prolific continent in the production of systematic review (**B**). The median global growth of systematic reviews was 26%

a median score of 0.71 (range 0.70–0.72) for 2020–2022 (Wilcoxon test, p-value < 0.001). Items with a notable increase over time were the reporting of a study protocol, and presentation of a flow chart for study inclusion. Before 2015, there was no mention of a study protocol in any SR. After 2015, there was an increase in systematic reviews reporting such a protocol with currently around 40% of systematic reviews fulfilling this criterium. None of the items showed a decrease.

The median quality score for the 5 most prolific countries were 0.73, 0.82, 0.73, 0.82, and 0.82 for the USA, China, UK, Brazil, and Iran, respectively. There was no statistically significant difference in the quality score between countries. All these countries show a quality increase over time (Fig. 6C). A quality score per country is presented in Supplementary Table 3.

Systematic reviews with a reported study protocol had higher quality scores compared to those without

such protocols (0.92, ± SD 0.12 versus 0.79±0.19, t-test: $p < 0.001$, excluding the item protocol from the analysis).

Discussion
Main findings
We aimed at summarizing the landscape of animal systematic reviews in neuroscience, including their global distribution, quality, and temporal trends. We found that (1) animal systematic reviews are increasingly employed, covering a broad range of topics, with many of them addressing translational research questions, (2) The most prolific countries in publishing systematic reviews are the USA, China, the UK, Brazil, and Iran, and (3) The quality of systematic reviews is generally high, increasing over time, and notably, there is no difference in quality between countries.
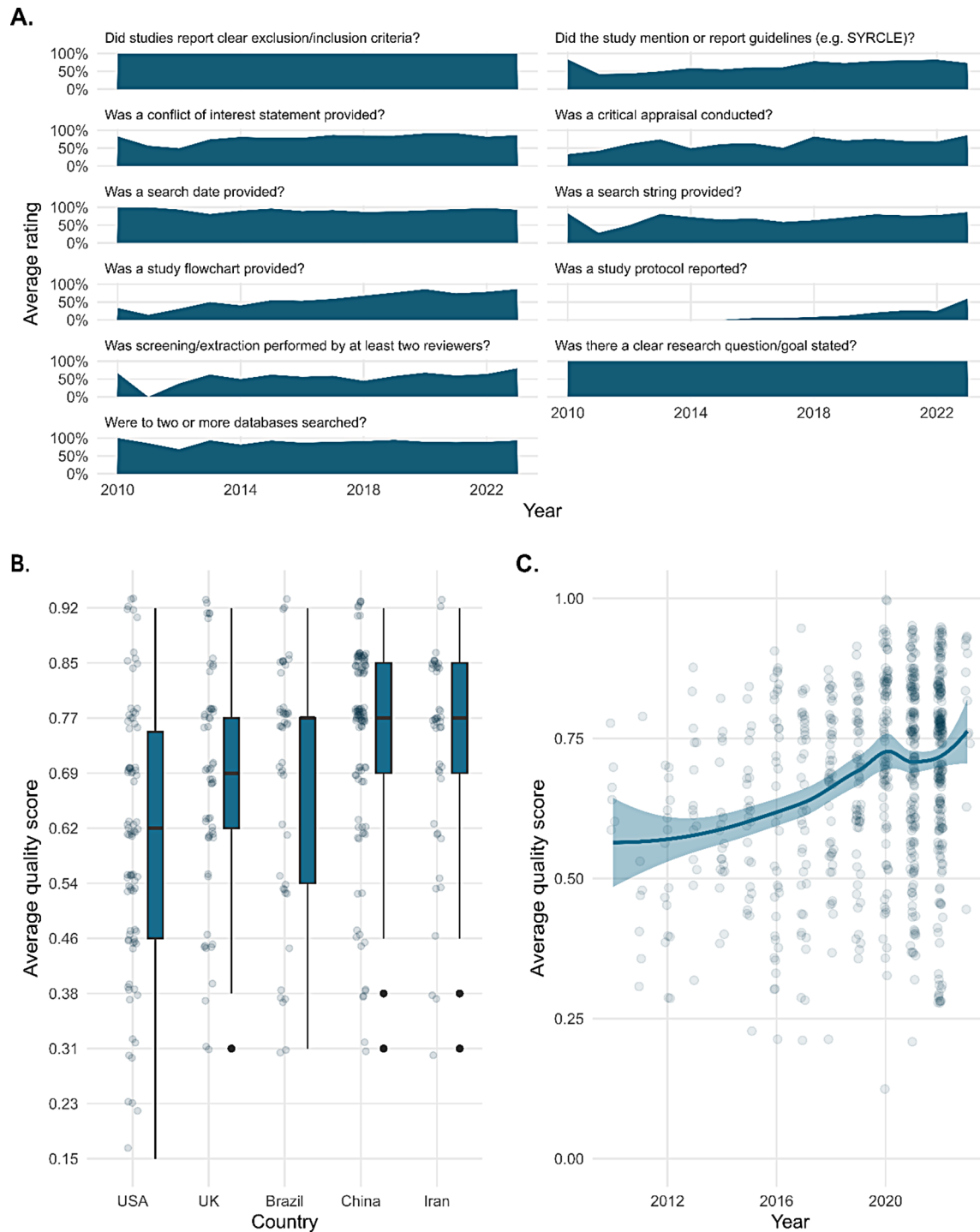
Fig. 6 Quality of animal systematic reviews. Quality of animal systematic reviews overall (**A**), per scored item (**B**), and for the most prolific countries publishing animal systematic reviews (**C**). There is a significant increase in systematic review quality over time. The countries show no statistically significant quality differences

Findings in the context of existing evidence

The increasing utilization of animal systematic reviews reflects a growing trend previously noted in older studies [30, 31]. This rise corresponds with the expanding publication of primary animal studies [18]. The value of systematic reviews, and meta-analyses where applicable, lies in their capacity to evaluate evidence while minimizing the risk of bias in the research review process. Their adoption also promotes transparency, assists in

identifying gaps in evidence, and aids in shaping future research directions and study designs.

In neuroscience, animal systematic reviews often focus on translational research, bridging animal and human studies. Stroke is the most frequently covered topic, despite its historically limited success in clinical translation [32, 33]. This focus is likely influenced by the early efforts of the CAMARADES consortium, which prioritized translational stroke research [9]. Other commonly reviewed topics include brain tumours and neurodegenerative diseases, which have also faced challenges in clinical translation [34, 35], as well as spinal cord injury, traumatic brain injury, and depression. Diseases with more successful translational outcomes, such as multiple sclerosis [8], epilepsy [36], and pain [37], are also represented in systematic reviews. Consistent with the translational emphasis of these reviews, most focus on potential therapeutic approaches, while those addressing diagnostic approaches remain relatively rare. This scarcity on diagnostics could reflect a general lack of robust animal models for diagnostic tool development, as these tools often require human-specific biomarkers and validation in clinical settings.

There are notable disparities in the coverage of diseases by systematic reviews, with some diseases being overrepresented and others showing clear gaps. Motor neuron diseases such as amyotrophic lateral sclerosis are relatively overrepresented when compared to their prevalence. This overrepresentation is likely driven in part by the high fatality rate of this relatively rare disease and the lack of available therapeutic options [38, 39]. Surprisingly, many common psychiatric diseases, including depression, bipolar disorders, autism spectrum disorder, eating disorders, attention deficit hyperactivity disorder, anxiety disorders, and addiction, are highly underrepresented, with only a few systematic reviews in comparison to their prevalence. One reason for this may be the perception that animal models are less reliable in mimicking complex behavioral phenotypes [40] or the high heterogeneity of underlying evidence, given the variety of animal models used to simulate specific psychiatric conditions (e.g., the chronic unpredictable mild stress model for depression and anxiety) [41]. Furthermore, several highly prevalent disorders, including obsessive-compulsive disorder, Down syndrome, meningitis, and tension-type headache, had no identified preclinical systematic reviews. This could reflect a lack of robust or widely accepted animal models for these conditions.

The most prolific publishers of animal systematic reviews are the USA, China, and the UK, aligning with their overall research output and funding [42, 43]. Interestingly, Brazil and Iran also emerge as leading contributors to animal systematic reviews, which contrasts with their lower ranking in the publication of systematic reviews in general [44]. This may be indicative of a stronger focus on preclinical research within these countries. Notably, Brazil's involvement with BRISA and the CAMARADES network likely supports their capability in this area.

The quality of systematic reviews has generally been high and has improved over time. Most reviews present clear research questions and define inclusion and exclusion criteria effectively. However, there is a notable gap in the critical appraisal of included studies and the publication of pre-study protocols. This aligns with earlier findings that placed the quality of animal systematic reviews between those dealing with in vitro and patient data [45]. Interestingly, we found no statistically significant quality differences across countries, consistent with recent comparisons of clinical systematic reviews from China and the USA [46]. This observation challenges previous perceptions of varying quality among meta-analyses, particularly from China [13, 47], suggesting that the adoption of formal reporting guidelines, like the preclinical PRISMA extension [23, 48], has positively impacted overall quality.

While animal systematic reviews can provide insights into disease mechanisms and translational research, we recognize the global shift towards reducing animal models in favor of new approach, or non-animal, methodologies (NAMs) [49, 50]. Nonetheless, systematic reviews of animal studies remain important due to the extensive body of existing animal research [18]. Additionally, systematic reviews can also incorporate data from non-animal evidence sources, making them a good tool to also address evidence from non-animal studies such as in vitro studies [3].

Limitations

Our study has certain limitations. First, our relatively broad definition of systematic reviews included some studies that do not strictly adhere to the Cochrane Collaboration's definition, with some reviews being in a more narrative fashion. This may lead to a slight underestimation of certain quality criteria; however, we believe this applies to only a small number of reviews. Second, we assessed quality based on reported information, which may not always accurately reflect the actual execution quality of the systematic reviews and meta-analyses. Finally, although we conducted a dual screening of the studies based on their abstracts, we did not perform a thorough examination of the full texts to confirm their eligibility for inclusion in our study. Despite this, we believe that the abstracts generally offer enough information to determine whether a study qualifies as a systematic review.

Strengths

Our study has the following strengths: First, with 1,358 included systematic reviews, it has a large sample size and thus offers a good foundation for analysis. Second, we developed and validated an automated tool for quality assessment, achieving high reliability and enabling efficient, large-scale evaluations. Third, our comprehensive mapping of topics and geographical trends provides insights into the focus areas and global distribution of animal research evidence synthesis.

Recommendations

Based on our findings, we call for the following two actions: First, an effort to strengthen preclinical evidence synthesis for psychiatric diseases, including depression, bipolar disorders, autism spectrum disorder, eating disorders, attention deficit hyperactivity disorder, and anxiety disorders but also for diseases like meningitis. Thus, efforts should be undertaken to communicate these findings to target stakeholders to improve implementation in practice, e.g., at conferences or scientific journals [51, 52]. Second, while the overall rigor of included systematic reviews and meta-analyses is relatively high, there is a need to emphasize pre-registration of study protocols and critical appraisal of included studies—two key strategies to mitigate common biases in systematic reviews [3]. This could be achieved by integrating these elements into teaching initiatives, such as summer schools [53] or online courses, and emphasizing them in methodological papers on systematic reviews.

Conclusions

With the rise of animal systematic reviews, and although generally of high quality, we identify concrete topical and quality targets to further enhance animal systematic reviews. Among them the recommended a priori publication of a study protocol. Such measures can contribute to the effective translation of animal research findings to clinical applications.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-05992-0.

Supplementary Material 1

Supplementary Material 2

Author contributions
Conception and design of study: BFH, DB, BVI. Acquisition of data: BFH, DB, SBKH, JB, BVI. Analysis of data: BFH, DB, MR, BVI. Drafting the initial manuscript: BFH, DB, MR, and BVI. All authors critically revised the paper draft.

Data availability
All data and code that support the findings of this study are available om the Open Science Framework at https://osf.io/wx5ta/. For any questions regarding data, meta-data, and analysis, contact the corresponding author BVI.

Declarations

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Ethical approval and consent to participate
Not applicable.

Author details
[1]Center for Reproducible Science, University of Zurich, Zurich, Switzerland
[2]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland
[3]Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland
[4]Department of Anaesthesiology, Pain and Palliative Medicine, Radboud university medical center, Nijmegen, The Netherlands
[5]Department of Neuroradiology, Clinical Neuroscience Center, University Hospital Zurich, Zurich, Switzerland

References
1. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane handbook for systematic reviews of interventions. Wiley; 2019.
2. Egger M, Higgins JP, Smith GD. Systematic reviews in health research: Meta-analysis in context. Wiley; 2022.
3. Ineichen BV, Held U, Salanti G, Macleod MR, Wever KE. Systematic review and meta-analysis of preclinical studies. Nat Reviews Methods Primers. 2024;4:72.
4. Ritskes-Hoitinga M, Pound P. The role of systematic reviews in identifying the limitations of preclinical animal research, 2000–2022: part 1. J R Soc Med. 2022;115:186–92.
5. Ioannidis JP. Systematic reviews for basic scientists: a different beast. Physiol Rev. 2022;103:1–5.
6. Hooijmans CR, Ritskes-Hoitinga M. Progress in using systematic reviews of animal studies to improve translational research. PLoS Med. 2013;10:e1001482.
7. de Vries RB, Wever KE, Avey MT, Stephens ML, Sena ES, Leenaars M. The usefulness of systematic reviews of animal experiments for the design of preclinical and clinical studies. Ilar j. 2014;55:427–37.
8. Hooijmans CR, Hlavica M, Schuler FAF, Good N, Good A, Baumgartner L, Galeno G, Schneider MP, Jung T, de Vries R, Ineichen BV. Remyelination promoting therapies in multiple sclerosis animal models: a systematic review and meta-analysis. Sci Rep. 2019;9:822.
9. Sena ES, Currie GL, McCann SK, Macleod MR, Howells DW. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. J Cereb Blood Flow Metabolism. 2014;34:737–42.
10. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, Macleod M, Mignini LE, Jayaram P, Khan KS. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ. 2007;334:197.

11. Besselink MG, van Santvoort HC, Buskens E, Boermeester MA, van Goor H, Timmerman HM, Nieuwenhuijs VB, Bollen TL, van Ramshorst B, Witteman BJ, et al. Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised, double-blind, placebo-controlled trial. Lancet. 2008;371:651–9.

12. Hooijmans CR, de Vries RB, Rovers MM, Gooszen HG, Ritskes-Hoitinga M. The effects of probiotic supplementation on experimental acute pancreatitis: a systematic review and meta-analysis. PLoS ONE. 2012;7:e48811.

13. Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q. 2016;94:485–514.

14. Hooijmans CR, Donders R, Magnuson K, Wever KE, Ergün M, Rooney AA, Walker V, Langendam MW. Assessment of key characteristics, methodology, and effect size measures used in meta-analysis of human-health-related animal studies. Res Synthesis Methods. 2022;13:790–806.

15. Langendam MW, Magnuson K, Williams AR, Walker VR, Howdeshell KL, Rooney AA, Hooijmans CR. Developing a database of systematic reviews of animal studies. Regul Toxicol Pharmacol. 2021;123:104940.

16. Hunniford VT, Montroy J, Fergusson DA, Avey MT, Wever KE, McCann SK, Foster M, Fox G, Lafreniere M, Ghaly M. Epidemiology and reporting characteristics of preclinical systematic reviews. PLoS Biol. 2021;19:e3001177.

17. van Luijk J, Bakker B, Rovers MM, Ritskes-Hoitinga M, de Vries RB, Leenaars M. Systematic reviews of animal studies; missing link in translational research? PLoS ONE. 2014;9:e89981.

18. Ineichen BV, Rosso M, Macleod MR. From data deluge to publomics: how AI can transform animal research. Lab Anim (NY). 2023;52:213–4.

19. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol. 2014;14:43.

20. Bahor Z, Liao J, Currie G, Ayder C, Macleod M, McCann SK, Bannach-Brown A, Wever K, Soliman N, Wang Q. Development and uptake of an online systematic review platform: the early years of the CAMARADES systematic review facility (SyRF). BMJ Open Sci. 2021;5:e100103.

21. Berg I, Härvelid P, Zürrer WE, Rosso M, Reich DS, Ineichen BV. Which experimental factors govern successful animal-to-human translation in multiple sclerosis drug development? A systematic review and meta-analysis. eBioMedicine 2024, 110.

22. Leenaars C, Tsaioun K, Stafleu F, Rooney K, Meijboom F, Ritskes-Hoitinga M, Bleich A. Reviewing the animal literature: how to describe and choose between different types of literature reviews. Lab Anim. 2021;55:129–41.

23. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Reviews. 2021;10:1–11.

24. de Vries RB, Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. Updated version of the Embase search filter for animal studies. Lab Anim. 2014;48:88.

25. Aromataris E, Fernandez RS, Godfrey C, Holly C, Khalil H, Tungpunkom P. Methodology for JBI umbrella reviews. 2014.

26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Reviews. 2016;5:210.

27. Ineichen BV, Furrer E, Grüninger SL, Zürrer WE, Macleod MR. Analysis of animal-to-human translation shows that only 5% of animal-tested therapeutic interventions obtain regulatory approval for human applications. PLoS Biol. 2024;22:e3002667.

28. Steinmetz JD, Seeher KM, Schiess N, Nichols E, Cao B, Servili C, Cavallera V, Cousin E, Hagins H, Moberg ME. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the global burden of Disease Study 2021. Lancet Neurol. 2024;23:344–81.

29. Collaborators GMD. 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. 2022;9:137–50.

30. Peters JL, Sutton AJ, Jones DR, Rushton L, Abrams KR. A systematic review of systematic reviews and meta-analyses of animal experiments with guidelines for reporting. J Environ Sci Health Part B. 2006;41:1245–58.

31. Korevaar D, Hooft L, Ter Riet G. Systematic reviews and meta-analyses of preclinical studies: publication bias in laboratory animal experiments. Lab Anim. 2011;45:225–30.

32. Campbell BC, De Silva DA, Macleod MR, Coutts SB, Schwamm LH, Davis SM, Donnan GA. Ischaemic stroke. Nat Reviews Disease Primers. 2019;5:1–22.

33. O'Collins VE, Macleod MR, Donnan GA, Horky LL, van der Worp BH, Howells DW. 1,026 experimental treatments in acute stroke. Ann Neurol. 2006;59:467–77.

34. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. Alzheimers Res Ther. 2014;6:1–7.

35. Hirst T, Vesterinen H, Conlin S, Egan KJ, Antonic A, Lawson McLean A, Macleod MR, Grant R, Brennan P, Sena E. A systematic review and meta-analysis of gene therapy in animal models of cerebral glioma: why did promise not translate to human therapy? Evidence-based Preclinical Med. 2014;1:21–33.

36. Regner GG, Pereira P, Leffa DT, de Oliveira C, Vercelino R, Fregni F, Torres IL. Preclinical to clinical translation of studies of transcranial direct-current stimulation in the treatment of epilepsy: a systematic review. Front NeuroSci. 2018;12:189.

37. Finnerup NB, Attal N, Haroutounian S, McNicol E, Baron R, Dworkin RH, Gilron I, Haanpaa M, Hansson P, Jensen TS, et al. Pharmacotherapy for neuropathic pain in adults: a systematic review and meta-analysis. Lancet Neurol. 2015;14:162–73.

38. Cannon AE, Zürrer WE, Zejlon C, Kulcsar Z, Lewandowski S, Piehl F, Granberg T, Ineichen BV. Neuroimaging findings in preclinical amyotrophic lateral sclerosis models—how well do they mimic the clinical phenotype? A systematic review. Front Veterinary Sci. 2023;10:1135282.

39. Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, Shaw PJ, Simmons Z, van den Berg LH. Amyotrophic lateral sclerosis. Nat Rev Dis Primers. 2017;3:17071.

40. Keehn J. Animal models for psychiatry. Routledge; 2018.

41. Antoniuk S, Bijata M, Ponimaskin E, Wlodarczyk J. Chronic unpredictable mild stress for modeling depression in rodents: Meta-analysis of model reliability. Neurosci Biobehavioral Reviews. 2019;99:101–16.

42. Docampo D, Bessoule J-J. A new approach to the analysis and evaluation of the research output of countries and institutions. Scientometrics. 2019;119:1207–25.

43. Wang X, Liu D, Ding K, Wang X. Science funding and research output: a study on 10 countries. Scientometrics. 2012;91:591–9.

44. Fontelo P, Liu F. A review of recent publication trends from top publishing countries. Syst Reviews. 2018;7:1–9.

45. Mignini LE, Khan KS. Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research. BMC Med Res Methodol. 2006;6:1–6.

46. Tian J, Zhang J, Ge L, Yang K, Song F. The methodological and reporting quality of systematic reviews from China and the USA are similar. J Clin Epidemiol. 2017;85:50–8.

47. Ioannidis JP, Chang CQ, Lam TK, Schully SD, Khoury MJ. The geometric increase in meta-analyses from China in the genomic era. PLoS ONE. 2013;8:e65602.

48. Lalu M, Moher D, Sena E, Macleod M, Avey M, Wever KE, Page MJ, McCann SK, Bohrer A, Sikora O et al. L, : PRISMA Extension for Preclinical In Vivo Animal Experiments. *OSF* 2020.

49. Kaplan BL, Hoberman AM, Slikker W, Smith MA, Corsini E, Knudsen TB, Marty MS, Sobrian SK, Fitzpatrick SC, Ratner MH. Protecting human and animal health: the road from animal models to new approach methods. Pharmacol Rev. 2024;76:251–66.

50. Taylor K, Modi S, Bailey J. An analysis of trends in the use of animal and non-animal methods in biomedical research and toxicology publications. Front Lab Chip Technol. 2024;3:1426895.

51. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, Proctor EK, Kirchner JE. A refined compilation of implementation strategies: results from the Expert recommendations for Implementing Change (ERIC) project. Implement Sci. 2015;10:1–14.

52. Waltz TJ, Powell BJ, Matthieu MM, Damschroder LJ, Chinman MJ, Smith JL, Proctor EK, Kirchner JE. Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert recommendations for Implementing Change (ERIC) study. Implement Sci. 2015;10:1–8.

53. Rosso M, Doneva SE, Howells DW, Leenaars CH, Ineichen BV. Summer school for systematic reviews of animal studies: fostering evidence-based and rigorous animal research. Altex. 2024;41:131–4.

Publisher's note