BMC Medical Research
Methodology

RESEARCH Open Access

CRIME-Q—a unifying tool for critical appraisal of methodological (technical) quality, quality of reporting and risk of bias in animal research

Mikkel Schou Andersen[1,2,3*], Mikkel Seremet Kofoed[1,2,3], Asger Sand Paludan-Müller[4,5,6], Christian Bonde Pedersen[1,2,3], Tiit Mathiesen[7,8,9,10], Christian Mawrin[11], Birgitte Brinkmann Olsen[2,12,13], Bo Halle[1,2,3] and Frantz Rom Poulsen[1,2,3]

Abstract

Background Systematic reviews within the field of animal research are becoming more common. However, in animal translational research, issues related to methodological quality and quality of reporting continue to arise, potentially leading to underestimation or overestimation of the effects of interventions or prevent studies from being replicated. The various tools and checklists available to ensure good-quality studies and proper reporting include both unique and/or overlapping items and/or simply lack necessary elements or are too situational to certain conditions or diseases. Currently, there is no tool available, which covers all aspects of animal models, from bench-top activities to animal facilities, hence a new tool is needed. This tool should be designed to be able to assess all kinds of animal studies such as old, new, low quality, high quality, interventional and noninterventional on. It should do this on multiple levels through items on quality of reporting, methodological (technical) quality, and risk of bias, for use in assessing the overall quality of studies involving animal research.

Methods During a systematic review of meningioma models in animals, we developed a novel unifying tool that can assess all types of animal studies from multiple perspectives. The tool was inspired by the Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies (CAMARADES) checklist, the ARRIVE 2.0 guidelines, and SYRCLE's risk of bias tool, while also incorporating unique items. We used the interrater agreement percentage and Cohen's kappa index to test the interrater agreement between two independent reviewers for the items in the tool.

Results There was high interrater agreement across all items (92.9%, 95% CI 91.0–94.8). Cohen's kappa index showed quality of reporting had the best mean index of 0.86 (95%-CI 0.78–0.94), methodological quality had a mean index of 0.83 (95%-CI 0.78–0.94) and finally the items from SYRCLE's risk of bias had a mean kappa index of 0.68 (95%-CI 0.57–0.79).

Conclusions The Critical Appraisal of Methodological (technical) Quality, Quality of Reporting and Risk of Bias in Animal Research (CRIME-Q) tool unifies a broad spectrum of information (both unique items and items inspired by other

*Correspondence:
Mikkel Schou Andersen
Mikkel.c.schou.andersen@rsyd.dk
Full list of author information is available at the end of the article

methods) about the quality of reporting and methodological (technical) quality, and contains items from SYRCLE's risk of bias. The tool is intended for use in assessing overall study quality across multiple domains and items and is not, unlike other tools, restricted to any particular model or study design (whether interventional or noninterventional). It is also easy to apply when designing and conducting animal experiments to ensure proper reporting and design in terms of replicability, transparency, and validity.

Keywords Animal research, Preclinical research, Preclinical studies, Systematic review, Preclinical methodology, Critical appraisal, Risk of bias, Methodological approach

Introduction

Systematic reviews and meta-analyses are at the pinnacle of evidence-based decisions [1]. Although systematic reviews of preclinical studies are not yet standard practice, the approach is becoming more common [2]. Systematic reviews within the field of animal studies should be performed with high methodological quality [3, 4], as they have the potential to be valuable in illuminating important gaps in translational science [5]. In accordance with the 3Rs (reduction, refinement, replacement), systematic reviews can integrate available observations and thereby synthesize a comprehensive view of preclinical knowledge for translation to clinical trials and to explicitly define pre-clinical knowledge gaps; a well-defined knowledge gap supports trial designs that avoid waste of animals and other resources as well as unnecessary duplication of effort by presenting relevant literature within a given field [6]. A major challenge for all systematic reviews is systematic violation of publication ethics and widespread availability of paper-mill papers and publications of inferior quality [7, 8]. Assessment of scientific quality is thus fundamental in the present publication landscape.

Several issues in the methodological quality and reporting of animal studies can lead to under- or overestimation of the effects of interventions or simply produce results that cannot be replicated [9]. These flaws can often be addressed through proper study design and reporting, aided by the implementation of the ARRIVE 2.0 guidelines for Animal Research: Reporting of In Vivo experiments [10], which is developed by the National Center of 3Rs. However, many studies still fail to adhere to these guidelines in reporting, and ultimately, quality remains insufficient [11]. Given that proper assessment of the quality studies included in systematic reviews is indispensable [12], various critical appraisal tools and checklists have been suggested for assessing study quality [10, 13–15]. The tools presented in the systematic review by Krauth et al. [13] describe different aspects of quality of reporting, methodological quality and risk of bias, and the elements/items for the various tools either overlap, are lacking, and/or are simply too situational to certain conditions/diseases [13]. Furthermore, there is a heavy focus on interventional rather than noninterventional studies for instance studies entailing the development of a specific animal model. Finally, the tools lack assessment of the laboratory work carried out in relation to an animal study. There are certain bench-top factors—such as cell handling prior to implantation or the mixing of compounds—that can influence the outcome, which is why it is crucial to assess them. Despite the many tools, none can be applied universally to preclinical studies and subsequently systematic reviews need apply different quality criteria for different studies in in one study.

In the absence of a validated tool for the critical assessment of records describing both in vitro and in vivo (and noninterventional) studies, we developed a critical appraisal tool, which fills the gaps in assessing the literature with new unique items, but is also inspired primarily by the following tools and checklists: Macleod et al. (2004) with the reliability and validity tested Collaborative Approach to Meta Analysis and Review of Animal Data from Experimental Studies (CAMARADES) checklist [16], Cramond et al. (2016) [14], the ARRIVE 2.0 [10] guidelines and the validated SYRCLE's risk of bias tool [15], which is based on the Cochrane Collaboration's tool for assessing risk of bias in randomized trials [17].

We have named the developed tool the Critical Appraisal of Methodological (technical) Quality, Quality of Reporting and Risk of Bias in Animal Research (or simply CRIME-Q). It addresses the quality of reporting, methodological quality (technical quality) and risk of bias and is applicable in potentially all kinds of animal studies. The quality of reporting refers to how well the information in the studies is described, ensuring that it can be replicated accurately. Methodological quality refers to the thoroughness and effectiveness of the technical aspects and performance of the studies, assessing whether the experiments were conducted with precision, and finally, risk of bias focuses on the various biases that can arise when designing and performing a study. The aims of the current study was to present and to discuss the CRIME-Q items, to display results from the internal validation through Cohen's kappa statistics and to describe how the CRIME-Q results can be presented.

Methods

Development

The CRIME-Q was developed during a systematic review of meningioma models in animals [18]. The author group's a priori knowledge of the literature in this area meant that a tool was needed that could assess both interventional and noninterventional results as well as assess studies on multiple levels in a nuanced manner in terms of quality of reporting, methodological (technical) quality and risk of bias and assess bench-top/laboratory work related to animal model. In the search of such a tool, we assessed multiple guidelines, checklists, and risk of bias tools for use in our systematic review but found that none of them covered the study area satisfactorily [10, 14–16, 19–44]. Hence, we found inspiration from and based the new tool on multiple sources, including and primarily the ARRIVE 2.0 [10], Macleod et al. (2004) with the CAMARADES checklist [16], Cramond et al. (2016) [14], and SYRCLE's Risk of Bias [15], and also included unique items not found in available tools. Inspiration was especially drawn from the ARRIVE 2.0 guidelines since these help ensure the optimal use of animals in research. The risk of bias (RoB) tool was included partly to complement quality of reporting (QoR) and methodological quality (MQ) but also to compare its performance against the QoR and MQ. The authors have distinct expertise to aid in the development of the tool within the fields of neurosurgery/oncology, epidemiology, and animal and cell-based models. Based on these experiences, items were presented to the author group for discussion. Some of the quality of reporting items were based on and inspired by other tools, as described, and other quality of reporting items were de novo (defined by the authors), such as 2X. The methodological quality (Y) categories were all de novo with the exception of sample size calculation. The risk of bias items were replicated from the existing tool [15]. The inclusion of specific items from Hoojimans et al.'s Syrcle's RoB tool [15] is deliberate, because it is a well validated tool. We chose the remaining items based on assessing beforementioned literature and through internal discussions. All the authors reviewed the final CRIME-Q items and agreed upon the definitions and final tool composition.

Primary sources of inspiration

As stated, the primary sources of inspiration were ARRIVE 2.0 [10] (the recommended), Macleod et al. (2004) with the CAMARADES checklist [16], Cramond et al. (2016) [14], and the SYRCLE risk of bias [15] tools. Table 1 provides an overview of the different items included in CRIME-Q and how they correspond with our primary sources of inspiration. Table 1 also shows, where CRIME-Q contains unique items, which is not present in any other available tools, we have assessed. The ARRIVE guidelines 2.0 are primarily used in reporting but can also be used in reviewing manuscripts. ARRIVE 2.0 assesses many of the desired and included items but does not contain information on the in vitro/bench-top part of the study, limitations, potential conflicts of interest, and publication bias. The CAMARADES checklist overlaps with ARRIVE 2.0 but also adds in terms of peer review and conflict of interest. It was developed for stroke models and contains items such as "blinded induction of ischemia" and "use of anesthetic without significant intrinsic neuroprotective activity," which cannot be extrapolated to reviewing literature in other areas. Cramond et al. describe the protocol for comparing the completeness of reporting of in vitro and in vivo research carried out in Nature Publication Group Journals when reviewing submitted manuscripts. In this publication, they included the signaling questions used in the appendix. None of the tool's 78 unique questions provide insight into any items on technical design aspects (handling of experiments; cells, animals, and experimental aspects) but give important and elaborate insight into statistical considerations (sample size and statistical plan), randomization, and blinding. Finally, SYRCLE's risk of bias was included because it is the validated risk of bias tool for animal studies and aligns with current literature.

Internal validation

Validation of CRIME-Q was performed via blinded assessment of all items of the tool by two reviewers (MSA and MSK). This was done during the preparation of a systematic review based on 114 published articles on animal models within the same field [18].

Statistical considerations

Cohen's Kappa (κ) statistics [45] was used to assess inter-rater reliability. We used the nominal Cohen's Kappa for all data (X, Y, Z). Cohen's Kappa was calculated using Stata/BE 17.0 (StataCorp LLC, TX, USA). Graphs and descriptive statistics were generated using GraphPad Prism 9.5.1 (GraphPad Software, MA, USA).

Description of CRIME-Q items and justification for their inclusion

The CRIME-Q tool is a multimodal critical appraisal tool for animal model studies with the intended purpose of providing an overview of a record or manuscript's overall validity and usability. Considering the heterogeneity (or simply lack thereof [11]) in the reporting of animal studies, the tool's items encompass the objective assessment of the quality of reporting – how well is the reporting done; methodological quality – how well were the study's

Table 1 Primary sources of inspiration for the CRIME-Q tool. This table displays the items from CRIME-Q, their sources of origin, and how these sources overlap or lack specific items

All CRIME-Q items		Type	CRIME-Q unique items	Arrive 2.0. The recommended [10]	The CAMARADES checklist [16]	Cramond et al. [14]	SYRCLE's risk of bias [15]
1X	Peer review	QoR			X		
2X	Bench-top/laboratory work related to establishing model—Reporting	QoR	X			(X)	
2Y	Bench-top/laboratory work related to establishing model – Methodology (technical quality)	MQ	X				
3X	Animals—Reporting	QoR		X	X	X	
3Y	Animals – Methodology (technical quality)	MQ	X				
3Z	Selection bias (baseline characteristics)	RoB		X			X
4Y	Sample size calculation	MQ		X	X	X	
5X	in vivo design and performance—Reporting	QoR		X	X		
5Y	in vivo design and performance – Methodology (technical)	MQ	X				
5Z (1)	Selection bias (Sequence generation)	RoB		X	X		X
5Z (2)	Performance bias (Random housing)	RoB		X			X
5Z (3)	Detection bias (Random outcome assessment)	RoB		X	X		X
6X	Compliance with animal welfare regulations	QoR		X	X		
7X	Blinding	QoR		X	X	X	X
7Z (1)	Performance bias (Blinding)	RoB		X	X	X	X
7Z (2)	Allocation bias (allocation concealment)	RoB		X		X	X
7Z (3)	Detection bias (blinding)	RoB		X	X	X	X
8X	Congruency between methods and results	QoR		X			
8Z (1)	Attrition bias (incomplete outcome data) (SYRCLE Item 8)	RoB					X
8Z (2)	Reporting bias (Selective outcome reporting) (SYRCLE Item 9)	RoB		X			X
9X	Presentation of limitations	QoR	X				
10X	Statement of potential conflict of interest	QoR			X		
10Z	Publication bias (influence) (SYRCLE Item 10)	RoB					X

QoR Quality of reporting, *MQ* Methodological quality, *RoB* Risk of Bias

technical aspects, choice of animals, proper handling of cells, and performance of experimentation/surgery executed; and risk of bias – assessing the design and execution of the study, from bench-top to in vivo experiments, which makes the tool unique.

The following section first describes the three main categories of the CRIME-Q: quality of reporting (QoR), methodological quality (MQ), and risk of bias (RoB). Furthermore, the tool is subdivided into domains (1–10) and 23 items. The 23 individual items are then described in turn with an indication of whether they relate to the QoR (X), MQ (Y), or RoB (Z). An overview is presented in Table 2.

Quality of reporting (QoR)

Transparency and replicability/reproducibility are central to scientific research [46]. Transparency ensures that findings and methods can be assessed by other researchers and helps to ensure an unbiased approach. Transparency requires that the reported results and methods used to obtain these results are traceable. Reproducibility provides credibility and means that research findings can be trusted with more certainty. Research is deemed reproducible when study results can be replicated using the same method and design.

The QoR refers to the quality of information that the records provide. The QoR is denoted as X in the following. Records should provide sufficient descriptions and details to ensure that studies can be replicated/reproduced. The information needed to assess the QoR will differ based on the specific topic. It is important for users of the CRIME-Q to make a predetermined list of information needed to replicate/reproduce the studies in question. The important items included should also be based on the 3Rs (reduction, refinement, replacement), which ensures the optimal use of animals in research.

Table 2 Critical Appraisal of Methodological (technical) Quality, Quality of Reporting and Risk of Bias in Animal Research (CRIME-Q) items, description, and potential biases/impacts

Items		Type	Questions and clarification	Potential impacts/biases
1X	Peer review	QoR	Did the paper undergo peer review prior to publication? Peer review might be useful for detecting errors or fraud. **Yes/No**	With peer review: Bias against negative studies Without: potential for errors and/or fraud
2X	Bench-top/laboratory work related to establishing model—reporting	QoR	Was the study's bench-top protocol sufficiently described (transparent, reproducible)? If e.g., cells are involved, did the study e.g., present incubator settings, description in detail of how the cells were treated (transfection, irradiation, etc.)? or describe how the cells were handled? Or for instance, did the study describe how to obtain a certain genetic model given there is no commercially available animal model? **Yes/Partly/No**	In vivo results are highly influenced by in vitro/bench-top part of the study. If not transparent and reproducible this lessens the usability of the study/model
2Y	Bench-top/laboratory work related to establishing model – methodology (technical)	MQ	Was the bench-top protocol feasible and technically well performed in relation to the experiment? Was it likely that the intended aim could be obtained based on the bench-top method? **Yes/Partly/No.** Studies with poor reporting (2X) will have difficulty gaining a high 2Y because of low transparency and the ability to assess method quality	If the bench-top protocol was not feasible, results may be misleading, and readers should use study/model with caution
3X	Animals—Reporting	QoR	Were the animals used in the experiment sufficiently described? Were all parameters: Type, breed, age, weight, and manufacturer sufficiently described? If type, age, and weight were sufficiently described OR if weight was missing, but the manufacturer was included a Yes was given. If only partly described, then Partly. and if we were not able to correctly identify the animals No was given. **Yes/Partly/No**	The importance of the description of animal type cannot be understated since the immunological profile differs from strain to strain, which influences results
3Y	Animals – Methodology (technical)	MQ	Did the study use similar baseline characteristics for the animals (age, weight, type)? And was the animals appropriate for the experiments (and appropriate strain)? **Yes/Partly/No.** Studies with poor reporting (3X) will have difficulty gaining a high 3Y because of low transparency and the ability to assess method quality	If animals were not homogenous or type is not appropriate for the study, study results might vary, e.g., low weight might result in poorer survival, which skews results. And xenografted material from e.g. human cannot be transferred to immunocompetent animals, hence they will fail to take
3Z	Selection bias (baseline characteristics) **(SYRCLE Item 2)**	RoB	Was the distribution of relevant baseline characteristics balanced between groups? I.e., was the distribution of e.g., male:female ratio, species, strain, age, and weight equally distributed throughout groups? **Yes/No/Unclear/NA**. Not applicable to studies using only one group	Unequal groups in intervention studies can skew results – introduces variables that potentially affect study results
4Y	Sample size calculation	MQ	Did the study include a calculation of sample size? Describe how it was calculated – at what power? Was it appropriate and well performed? **Yes/Partly/No**	Studies may prove under/overpowered in terms of drug efficacy if too few/many animals were used
5X	in vivo design and performance—Reporting	QoR	Description of the in vivo study part: Were the surgery, implantation/injection method, and duration (whole experiment) sufficiently described? Is the study transparent and reproducible? **Yes/Partly/No**	The results will be difficult to replicated, if the study is poorly described. Meaning the study is difficult to properly be assessed as a useful base for further research

Table 2 (continued)

Items		Type	Questions and clarification	Potential impacts/biases
5Y	in vivo design and performance -Methodology (technical)	MQ	Did the method seem feasible and technically well performed concerning the study's aim and outcome and in contrast to other known literature? Is it likely that the in vivo study design influences the results—incomprehensive/insensible method? **Yes/Partly/No**. Studies with poor reporting (5X) will have difficulty gaining a high 5Y because of low transparency and the ability to assess method quality	A poor methodology can skew results making conclusions in relation to aims obsolete
5Z (1)	Selection bias (Sequence generation) **(SYRCLE Item 1)**	RoB	Was there a description of allocation (the process by which experimental units are assigned to experimental groups)—And was it appropriate? Not applicable to nonintervention studies. **Yes/No/Unclear/NA**	Unequal groups in intervention studies can skew results – introduces variables that potentially affect study results
5Z (2)	Performance bias (Random housing) **(SYRCLE Item 4)**	RoB	Were the animals randomly housed during the experiment? Yes/no/unclear. Not applicable to nonintervention studies. **Yes/No/Unclear/NA**	Some types of experiments are influenced by the location of housing, hence random assignment of placement could negate these issues
5Z (3)	Detection bias (Random outcome assessment) **(SYRCLE Item 6)**	RoB	Were animals randomly selected for outcome? For instance, If human endpoints (i.e., poor conditions, weight, etc.) were met and the investigators were not blinded, then the outcome cannot be assessed randomly. Not applicable to nonintervention studies. **Yes/No/Unclear/NA**	Bias toward assessing intervention effect size
6X	Compliance with animal welfare regulations	QoR	Did the study comply with any animal welfare regulations? **Yes/Partly/No**	Assurance of proper animal care throughout the study. Also important in terms of survival studies (human endpoints vs. death)
7X	Blinding	QoR	Was the study blinded in any way? Was the outcome assessed in a blinded fashion? Were the animals randomly selected across all groups of e.g., intervention? Were the investigator or animal handlers blinded? **Yes/Partly/No** More specific blinding is listed below in 7Z(1–3)	Untranslatable results to human conditions. Blinding is a strategy for reducing the risk that researchers, animal care staff, and others involved may influence outcomes (subconsciously or otherwise)
7Z (1)	Performance bias (Blinding) **(SYRCLE Item 5)**	RoB	Describe all used means, if any, to blind trial caregiver and researchers from knowing which intervention each animal received. **Yes/No/Unclear/NA**. Not applicable for nonintervention studies, however, it could be applicable for instance in xenograft studies, where multiple patient samples were used	Animal handling may be affected by unblinded study design
7Z (2)	Allocation bias (allocation concealment) **(SYRCLE Item 3)**	RoB	Could the investigator allocating the animals to intervention or control group not foresee assignment? **Yes/No/Unclear/NA.** Not applicable to nonintervention studies. Yes/no/unclear/NA. This could be applicable for instance in xenograft studies, where multiple patient samples were used	In relation to 7Z (1). Selection, handling, and treatment of animals may be affected if allocation concealment was not adequately performed

Table 2 (continued)

Items		Type	Questions and clarification	Potential impacts/biases
7Z (3)	Detection bias (blinding) **(SYRCLE Item 7)**	RoB	Was the outcome assessor blinded? and could the blinding have been broken? Describe all measures used, if any, to blind outcome assessors from knowing which intervention each animal received. Were the outcome assessment methods the same in each group? **Yes/No/Unclear/NA**. This could be applicable to instance in xenograft studies, where multiple patient samples where used	Measurement of the outcome can be over/underestimated if proper blinded outcome assessment was not performed
8X	Congruency between methods and results	QoR	Did the study present all their findings based on the methods described? Is there congruency between the method and results sections? And is it transparent? **Yes/Partly/No**	Presenting results in which methods are not described is not transparent and replicable and should be interpreted with caution
8Z (1)	Attrition bias (incomplete outcome data) **(SYRCLE Item 8)**	RoB	Describe the completeness of outcome data including attrition and exclusions from the analysis and were incomplete outcome data adequately described? Were all animals included in the analysis and if not, was it described why they were not included? **Yes/No/Unclear**	Attritions and/or exclusions should be clearly described, i.e., the number of animals used. If not, study results become difficult to assess. Poor replicability and transparency
8Z (2)	Reporting bias (Selective outcome reporting) **(SYRCLE Item 9)**	RoB	Was the study protocol available (require a description of protocol location in the article) and were all of the study's prespecified primary and secondary outcomes reported in the manuscript? Was the study protocol not available but was it clear that the published report included all expected outcomes (i.e., comparing methods and results sections)? The study report fails to include results for a key outcome that would be expected to have been reported for such a study, i.e., tumor-take rate in transplantation experiments. **Yes/No/Unclear**	Congruency between results and methods should be carefully described to avoid reporting bias. If key outcomes for a certain method were not described, study validity, transparency, and replicability become difficult
9X	Presentation of limitations	QoR	Did the study contain a section of limitations, or did they comment on the limitations of the study in relationship to in vitro and/or in vivo subparts? **Yes/Partly/No**	No study is without limitations, and it is paramount to present them to the reader for transparency's sake
10X	Statement of potential conflict of interest	QoR	Did the study contain a statement of potential conflicts of interest? **Yes/No**	Potential conflicts of interest can skew results, i.e., if an investigator has a method patent or is paid by a certain pharmaceutical company, hence it is important for transparency's sake to include it in the study
10Z	Publication bias (influence) (**SYRCLE Item 10)**	RoB	Inappropriate influence of funders or biased by companies. Was the study free of inappropriate influence from funders or companies supplying drugs or equipment? Did the authors declare a direct conflict of interest in relation to the study? Yes: Conflict of interest statement with no conflicts of interest. **Yes/No/Unclear**	Publication bias – Negative results will be less likely to be published if inappropriate influence of funder or biased companies occur

QoR Quality of Reporting, *MQ* Methodological Quality, *RoB* Risk of Bias

Items from ARRIVE 2.0 was used for this purpose as primary source of inspiration. Further inspiration for QoR items in relation to the 3Rs can be found in Hooijmans et al. 2010 [24]. If the study is fully described in accordance with the individual QoR items, then a full "X" is recorded. If the study is only partially described (i.e., some details are missing in accordance to the predetermined definition but the study can be reproduced at least partially), then "(X)" is recorded. If studies are seriously lacking in descriptions and information, a "0" is recorded.

Methodological quality (MQ)

To assess methodological quality (technical quality), the reviewers need to be knowledgeable within the field covered by the studies. Methodological quality in the CRIME-Q tool is an assessment of how feasible the results are, based on the methods presented and whether the study was well performed technically. Could the results be achieved based on the technical performance of the experiments? For instance, did the researchers handle the specific tissue or cells appropriately and show correct use of culture medium, incubator settings, and surgical methods, as well as address any issues related to, e.g., a wide age range or weight gaps in the animals. A fundamental aspect is the use of controls—positive and negative controls must be included for all analyses where false positive or negative findings are possible. This relates to both laboratory techniques and animal experiments.

It is clear that poor reporting quality affects the assessment of methodological quality. In the CRIME-Q tool, MQ only applies to the key elements of bench-top, animals, and in vivo design/performance. MQ in CRIME-Q is dependent on the assessor group's knowledge within a given field, meaning the definition of a well-performed method/study will be based on the author group's definitions. It is important to note that a predetermined definition of quality is needed prior to assessments, e.g., a description of proper handling of a certain tissue or the definition of an appropriate animal type, e.g., the mice used for neurobehavioral studies would be suboptimal to the superior rat in this regard. If a study is considered to show good (technical) performance, a "Y" is recorded. If it shows partially good performance, a "(Y)" is recorded, and a study that is poorly performed/cannot be assessed is recorded as "0".

Risk of bias (RoB)

An assessment of risk of bias (RoB) is essential in systematic reviews. Its wide use in clinical systematic reviews has led to the Cochrane Collaboration developing a validated tool—the Cochrane Risk of Bias Tool [17, 47]. The ability of a systematic review to draw reliable conclusions relies on the validity of the data and the results/conclusions of the included studies. Many studies have issues in this regard—more so for animal studies, which show severe deficits [26]. SYRCLE's RoB tool [15] is a validated tool to assess animal intervention studies. It contains ten elements and was based on the Cochrane RoB tool, with which it shares five items. We find it crucial to include elements from SYRCLE's RoB tool in the CRIME-Q to ensure uniformity and comparability to previous work that uses SYRCLE's RoB tool. Furthermore, items of SYRCLE's RoB tool have been placed in appropriate domains, i.e., 3. Animals, 5. In vivo design and performance, 7. Blinding, 8. Congruency in the data and methods and 10. Statement of potential conflicts of interest for easier assessment and overview of the individual domains. It is important to note that some of the SYRCLE's RoB elements are applicable only in intervention studies; these items are recorded as "not applicable/NA" when using the CRIME-Q. Otherwise, the RoB assessment is recorded as yes, no, or unclear. Furthermore, SYRCLE's risk of bias item 10 contains more examples under the overall question of "Was the study apparently free of other problems that could result in high risk of bias?", which could be included if needed based on the area of interest. For item 10 we have chosen to include publication bias (influence), since it is the broadest and most applicable generally.

Peer review

Quality of reporting

Peer review is the standard method for selecting scientific work for publication and improving the quality of research, where peer reviewers' comments identify manuscript flaws [48]. It is important to note that peer review also results in bias against negative studies [49]. There are limitations to the peer review process and it is prone to inconsistencies in the assessments [50]. It is important to note not peer reviewed studies can be of good quality. The inclusion of peer review in this tool is to provide the user and reader with transparency regarding the assessed studies. Peer review has been included in this critical appraisal tool and is categorized as quality of reporting (QoR). It can be assessed via the question: Did the paper undergo peer review prior to publication? If the study was peer reviewed, record "X"; and if not record "0".

Bench-top activities

Quality of reporting This item refers to all bench-top/laboratory work carried out in relation to the animal model described in a given study. The most important aspect is the ability of the available information to produce replicable results. Information required to replicate experiments could entail the following. For example, if working with cells/tissues: i) Cell line description and origin (how/where were the established cell lines acquired, or type of tumor/patient regarding primary cells), is the passage number described (if primary cells), and how were the cells/tissues handled; in what culture medium were the cells grown in and resuspended in. ii) What were the incubator settings? iii) Which reagents were used; description of medium including add-ons (fetal bovine serum percentage and other reagents) and origin (and lot number if appropriate). iv) Was there any testing of pathogens in the laboratory e.g., mycoplasma. Or, for example, if a genetically engineered model was used, how was it created, and which system was used? Basically, all aspects should be described to such a degree that the model can be replicated. This can differ depending on the kind of model being assessed. Authors of systematic animal study reviews should be clear as to which aspects were assessed. If the study is deemed fully replicable, record "X"; if it is only partially replicable, record "(X); if it is deemed unreproducible, record "0".

Methodological quality The methodological approach or quality of the bench-top/laboratory practice is based on how well was performed and whether the results and method were feasible with the method presented. There are certain bench-top factors that can influence the outcome. While the animal model and experimental aspects might be sound, suboptimal bench-top procedures prior to in vivo experiments can still affect results. For instance, were the incubator settings appropriate for the type of cells? Were reagents handled appropriately? In general, was the bench-top protocol feasible and well performed in relation to the in vivo experiment? Was it likely that the intended aim could be obtained based on the bench-top method? If yes, then record "Y"; if in doubt, but there are good elements, record "(Y)"; if it is deemed insufficient, then record "0". Studies with poor reporting (2X) will have difficulty obtaining high 2Y because of low transparency and limited ability to assess methodological quality.

Animals

Quality of reporting A proper description of the animals used in a study is crucial, as the results can be directly linked to certain species, strains, and even sex. Were the animals used in the experiment sufficiently described? Were all the animal parameters (species, strain, age, weight, sex, and manufacturer) sufficiently described as indicated in the ARRIVE 2.0 guidelines [10]. A full "X" is given if the species, strain, age, and weight are sufficiently described or if just the weight is missing but the manufacturer is included (minor details missing). If only partially described, then "(X)" is recorded. If the user is not able to correctly identify the animals, the score is recorded as "0".

Methodological quality This section can be answered with the simple question: Were the animals used in vivo appropriate for the experiments? e.g., were immunodeficient or immunosuppressed animals used for xenografts; Were there any incongruencies or broad baseline characteristic ranges (e.g., in age, weight) that were not addressed and explained? If the animals chosen are appropriate for the study, then record "Y"; if in doubt (due to type or description), but is probably appropriate, record "(Y)"; if the animals are deemed not appropriate, then record "0". Studies with poor reporting (3X) will have difficulty obtaining high 3Y because of low transparency and limited ability to assess methodological quality.

Risk of bias Selection bias (baseline characteristics) (Item 2 in [15]) Was the distribution of relevant baseline characteristics balanced between groups (i.e., was the distribution of, e.g., sex or weight equally distributed throughout groups)? Did the study use similar baseline characteristics for the animals (species, strain, age, weight, type)? Yes/no/unclear/NA. Given that this item is applicable only for intervention studies, nonintervention studies will be recorded as "not applicable/NA".

Sample size calculation

Methodological quality Although sample size is an important aspect of research, it is often not considered in experimental in vivo designs [51]. An underpowered study might miss any significant differences, while an overpowered study might waste resources [52]. There are two methods for calculating sample size: calculating sample size based on power analysis (recommended) or using a resource equation, which should be used when it is not possible to assume effect size [37]. The sample size calculation is of methodological quality origin and hence is included here. Questions to be asked include: Does the study state how the sample size was chosen to ensure adequate power to detect effect size a priori, and/or is there a statement regarding sample size even if no statistical methods were used? If records include sample size (or a statement if no sample size) and it is appropriate, then record"Yes"/"Y", if only partly described"Partly"/"(Y)", and if it no description or inappropriate use then"No"/0.

In vivo design and performance

Quality of reporting A description of the in vivo design and performance and the determination of transparency and replicability based on the information given can be complicated. Useful descriptions are given in the "Essential 10 and Recommended" items of the ARRIVE 2.0 guidelines [10] and are utilized in this section. In general, a description should be as detailed as possible for a third party to perform and replicate results in terms of study design, description of the experiment and how it was performed, description of groups, duration of intervention, and which outcome measurements were used to assess desired outcomes. Failure to disclose all outcome measurements chosen before the start of the study introduces positive outcome bias, where only significant positive outcomes are reported in contrast to negative outcomes [53, 54]. It is paramount for users of the CRIME-Q tool to describe which aspects were used to assess the quality of reporting for the specific model under review with a predetermined list of information needed.

Some possible examples are as follows:

- Was the surgery/procedure to establish the model sufficiently described? Design, performance, sham surgery, etc.
- Was the duration/time frame/follow-up time of the experiment sufficiently described?
- Was the implantation process described in detail? Number of cells and buffer solution used, place of injection with coordinates, use of a stereotactic frame or not, description of anesthesia (dose and route of administration), and how euthanasia was performed?
- Intervention: was the randomization process described (random allocation treatment)?
- Is the number of animals reported? (if animals were discarded, was this described?)
- Use of inclusion/exclusion criteria when selecting animals?
- A thorough description of outcome measures, i.e., size of tumor after treatment, behavioral changes, molecular markers, etc., and were the primary and secondary outcomes adequately described?
- How were the animal experimental groups kept, e.g., single animals, litter, cages?

Methodological quality Did the methods/experiments technically seem feasible and well performed with respect to the study's aim and outcome in contrast to/in line with known literature in the field? Is it likely that the in vivo design or performance influenced the results due to incomprehensive or inappropriate methods? For example, was the surgery well-performed based on the description, or was it plausible that the duration of the experiment would align with the expected pharmaceutical response? Additionally, are the statistics reported in such a way that variability and risk of type-2 error can be assessed, for instance through the use of 95% confidence intervals? If the study is deemed of high methodological (technical) quality, then record "Y"; if in doubt, but there are good elements, record "(Y)"; if it is deemed insufficient or poor, then record "0". Studies with poor reporting (5X) will have difficulty obtaining high 5Y because of low transparency and the limited ability to assess methodological quality.

Risk of bias

- Allocation bias (Sequence generation) (Item 1 in [15]) Was there a description of allocation (the process by which experimental units are assigned to experimental groups)? Yes/no/unclear/NA. Given that this item is applicable only to intervention studies, nonintervention studies will be recorded as "not applicable/NA". For a more lenient assessment, we recorded yes ("Z") if randomization was present.
- Performance bias (Random housing) (Item 4 in [15]) Were the animals randomly housed during the experiment? Yes/no/unclear/NA. This item is applicable only to intervention studies; hence, nonintervention studies will be noted as 'not applicable/NA.
- Detection bias (Random outcome assessment) (Item 6 in [15]) Were animals randomly selected for the outcome? For instance, if human endpoints (e.g., poor condition/poor health state, weight, etc.) were met and led to competing endpoints (some animals reached the primary endpoint and others met human endpoints; hence, terminating the experiment prematurely) and if the investigators were not blinded, then the outcome could not be assessed randomly. The groups could be described as having competing endpoints. Yes/no/unclear/NA. This item is applicable only to intervention studies; hence, nonintervention studies will be noted as 'not applicable/NA'.

Compliance with animal welfare regulations

Quality of reporting The ethical concerns associated with animal research differ between countries. A common directive has been implemented in Europe that defines minimum standards for the use of animals, where the 3Rs (reduction, refinement, replacement) should be taken into consideration [55]. Compliance with the 3Rs is of utmost importance in all animal research and should always be considered first when designing and performing animal studies. Hence, a statement of compliance with animal welfare regulations in regard to animal ethics is important. This item may not influence the validity of the results if it is not met. In particular, older studies do not present such statements, however it is expected in newer studies. This item can be answered with the following question: Did the study comply with animal welfare regulations, or did the authors state specific ethical considerations about animal welfare? If yes, record “X”; if no or no statement, record “0”.

Blinding

Quality of reporting In vivo experiments depend on unbiased results for proper translation to humans. Blinding is a strategy for reducing the risk that researchers, animal care staff, and others involved may influence outcomes (subconsciously or otherwise). A lack of blinding can lead to overestimation of the results and to false positive findings [56]. Regarding the quality of reporting, we have decided to include whether a given study includes any form of blinding and, if described, what type of blinding. If described, then record “X”(yes); if unclearly described but blinding exists in some capacity, record “(X)”; if not described, record “0”(no). The risk of bias section below covers how the blinding is utilized and where it is applied in the studies.

Risk of bias

- Performance bias (Blinding) (Item 5 in [15]) Describe all the means used, if any, to blind trial caregivers and researchers from knowing which intervention each animal received. Yes/no/unclear/NA. This item is applicable only to intervention studies; hence, nonintervention studies will be noted as ‘not applicable/NA’).
- Allocation bias (Allocation concealment) (Item 3 in [15]) Could the investigator allocating the animals to the intervention or control group not foresee assignment? Yes/no/unclear/NA. This item is applicable only to intervention studies; hence, nonintervention studies will be noted as ‘not applicable/NA’.
- Detection bias (Blinding) (Item 7 in [15]) Was the outcome assessor blinded? Could the blinding have been broken? Describe all measures used, if any, to blind outcome assessors from knowing which intervention each animal received. Were the outcome assessment methods the same in both groups? Yes/no/unclear/NA. This item is applicable only to intervention studies; hence, nonintervention studies will be noted as ‘not applicable/NA’).

Congruency in data and methods

Quality of reporting This section describes the laboratory methods and outcome data used and the congruency between them. For instance, if the study used immunohistochemistry (e.g., antibodies, lot numbers), are the data shown based on what is described in the methods? Other examples include scan settings and type, and other methods such as western blot, RNA-sequencing, DNA methylation, etc. Basically, did the study present all their findings based on the methods described? Was there congruency between the methods and results? Was it all described—which data were not included? *Did the study present methods and data/results in a transparent manner?* If there is congruency in methods and data and is it transparent, then record “X”(yes); if unclearly described but it is likely, record “(X)”; if not described, record “0”(no).

Risk of bias

- Attrition bias (incomplete outcome data) (Item 8 in [15]) Describe the completeness of the outcome data, including attrition and exclusions, from the analysis. Were incomplete outcome data adequately described? Or another example: were all animals included in the analysis and if not, was it explained why they were not included? Yes/no/unclear.
- Reporting bias (Selective outcome reporting) (Item 9 in [15]) Was the study protocol available (requires a description of the protocol location in paper), e.g. is the protocol available at https://preclinicaltrials.eu/, https://www.animalstudyregistry.org/, open science framework (https://osf.io/), and were all of the study’s prespecified primary and secondary outcomes reported in the manuscript? If the study protocol was not available, was it clear that the published report included all expected outcomes (i.e., comparing the Methods and Results sections)? Does the study report fail to include results for a key outcome that would be expected to have been reported for such a study e.g., tumor take rate in transplantation experiments? Yes/no/unclear. Of note, it is crucial to preregister studies to maintain complete transparency. Failing to distinguish between generating post-dictions and

testing pre-dictions can undermine the credibility of research findings. However, inherent biases in human reasoning, such as hindsight bias, make it challenging to prevent this error [57].

Presentation of limitations

Quality of reporting Presenting the limitations of the experiment is crucial. Limitations represent weaknesses in a given study that may influence conclusions and outcomes. The objective of presenting limitations is to ensure transparency and discuss/present the drawbacks of a study to the readers. Too often, the limitations of scientific works are overlooked or reduced to simplistic and minimally relevant themes [58]. Did the study contain a section of limitations, or did the authors comment on the study limitations in relation to in vitro and/or in vivo subparts? If done so in a satisfactory manner, record "X"; if the authors discuss limitations but do not go into sufficient depth, record "(X)"; if no limitations were presented, record "0".

Statement of potential conflict of interest

Quality of reporting The primary goal of all research is to provide research the community with impartial unbiased evidence, but several other factors can affect results, such as creating a profit or benefitting a career [59]. Potential conflicts of interest can skew the results, e.g., if an investigator has a method patent or is paid by a pharmaceutical company. Hence, it is important for transparency's sake to include all such potential conflicts. This section can be answered with the question: Did the study contain a statement of potential conflict of interest? If yes, record "X"; if no, record "0". Yes/no.

An additional note is that there is a risk of publisher conflict of interest in regard to if the authors supported publication by paying a processing fee and if the publishers were aware that an article processing fee would be paid when deciding to accept the paper. However, this can be more difficult to assess, especially in older research; hence, this item is not included but important to acknowledge.

Risk of bias Other bias (inappropriate influence of funders or biased by companies) (Item 10 in [15]). Was the study free of inappropriate influence from funders or companies supplying drugs or equipment? (Did the authors declare a direct conflict of interest to the study?) "Z" (yes): Conflict of interest statement with no conflict of interests. "0" (no): Conflict of interest statement with conflict of interest/influence of funders. And if unclear this should be recorded. Yes/no/unclear. It is also important to note that there are situations where other risks of bias should be included. SYRCLE's RoB includes *"Was the study apparently free of other problems that could result in high risk of bias?"*, where funding is only one of them.

CRIME-Q instructions, presentation, performance, internal validation and interpretation

Instructions

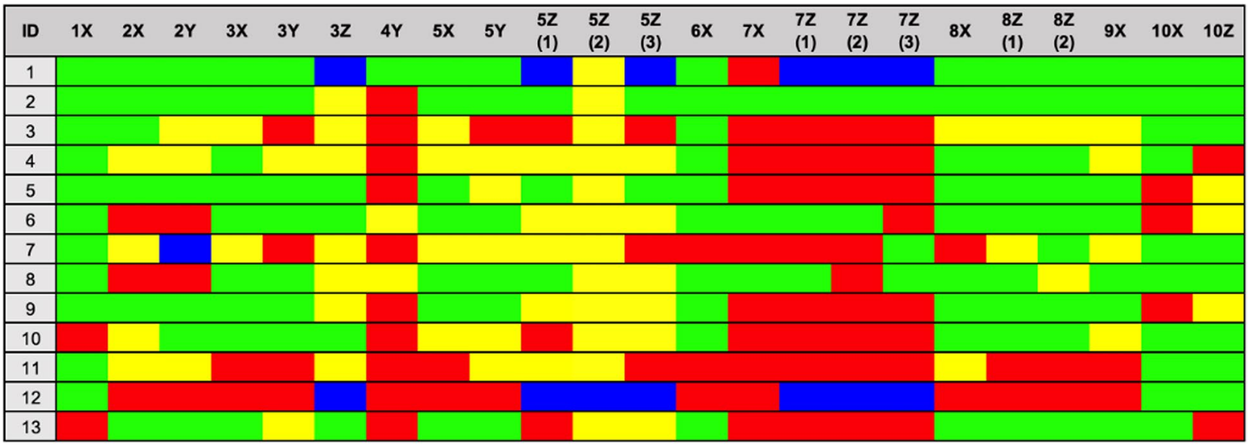
Before using the CRIME-Q to assess the overall quality of reporting and methodological quality (including risk of bias based on items from SYRCLE's RoB tool), it is important to create a predetermined list of items to assess. This predetermined list of items/information represents items in each category that need to be present/absent for identifying the answer categories (yes/no/partly/unclear/not applicable). This list should always be included as supplemental material for the sake of transparency and replicability. An example of such a list (for assessing tumor models) is available in Additional File 1. Assessment of risk of bias can be performed using signal questions from SYRCLE's RoB tool [15]—all deviations should be noted in the same additional file mentioned above. At least two reviewers should independently apply CRIME-Q while being blinded to each other's assessments. All incongruencies in answers for any given category should be discussed between the assessors—and possibly with a third team member if an agreement cannot be reached.

Presentation of results

We offer two ways of presenting the results of CRIME-Q, depending on the number of included records. Presentation in a colored grid (Fig. 1) is visually appealing and provides a transparent presentation of the results. Such a grid table becomes too large when there are many records (e.g. > 40); hence another approach could be more useful as shown in Fig. 2. Here, the overall percentages for each category can be assessed for the overall critical appraisal—e.g., divided into X and Y categories in Fig. 2A and Z categories in Fig. 2B. If this solution is chosen, the full assessment table should be attached as supplemental material.

Interrater reliability – performance and validation

In our systematic review of meningioma animal models [18], two reviewers used CRIME-Q to score all the unique 114 records in a blinded fashion. We subjected the blinded assessments to Cohen's Kappa. Final scoring for the systematic review was done in consensus and all disagreements were handled through discussions and/or senior author if necessary. The interpretation of the



ID	1X	2X	2Y	3X	3Y	3Z	4Y	5X	5Y	5Z (1)	5Z (2)	5Z (3)	6X	7X	7Z (1)	7Z (2)	7Z (3)	8X	8Z (1)	8Z (2)	9X	10X	10Z
1	🟩	🟩	🟩	🟩	🟩	🟦	🟩	🟩	🟩	🟦	🟨	🟦	🟩	🟥	🟦	🟦	🟦	🟩	🟩	🟩	🟩	🟩	🟩
2	🟩	🟩	🟩	🟩	🟩	🟨	🟥	🟩	🟩	🟩	🟨	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
3	🟩	🟩	🟨	🟨	🟥	🟨	🟥	🟨	🟥	🟥	🟨	🟥	🟩	🟥	🟥	🟥	🟥	🟨	🟨	🟨	🟨	🟩	🟩
4	🟩	🟨	🟨	🟩	🟨	🟨	🟥	🟨	🟨	🟨	🟨	🟨	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟨	🟩	🟥
5	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟨	🟩	🟨	🟩	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟥	🟨
6	🟩	🟥	🟥	🟩	🟩	🟩	🟨	🟩	🟩	🟨	🟨	🟨	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟥	🟨
7	🟩	🟨	🟦	🟨	🟥	🟨	🟥	🟨	🟨	🟨	🟨	🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟨	🟩	🟨	🟩	🟩
8	🟩	🟥	🟥	🟩	🟩	🟨	🟨	🟩	🟩	🟩	🟨	🟨	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟨	🟩	🟩	🟩
9	🟩	🟩	🟩	🟩	🟩	🟨	🟥	🟩	🟩	🟨	🟨	🟨	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟥	🟨
10	🟥	🟨	🟩	🟩	🟩	🟩	🟥	🟨	🟨	🟥	🟨	🟨	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟨	🟩	🟩
11	🟩	🟨	🟨	🟥	🟥	🟨	🟥	🟥	🟨	🟨	🟨	🟥	🟥	🟥	🟥	🟥	🟥	🟨	🟥	🟥	🟥	🟩	🟩
12	🟩	🟥	🟥	🟥	🟥	🟦	🟥	🟥	🟥	🟦	🟦	🟦	🟥	🟥	🟦	🟦	🟦	🟥	🟥	🟥	🟥	🟩	🟩
13	🟥	🟩	🟩	🟩	🟨	🟩	🟥	🟩	🟩	🟥	🟨	🟨	🟩	🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟩	🟥

Fig. 1 Presentation example of few included studies. Green: Yes, Yellow: Partly/unclear, Red: No, Blue: NA. Fictitious data 1X Peer review, 2X Bench-top, 2Y Bench-top, 3X Animals, 3Y Animals, 3Z Selection bias (baseline characteristics), 4Y Sample size, 5X in vivo design and performance, 5Y in vivo design and performance, 5Z (1) Selection bias (Sequence generation), 5Z (2) Performance bias (Random housing), 5Z (3) Detection bias (Random outcome assessment), 6X Animal welfare compliance, 7X Blinding, 7Z (1) Performance bias (Blinding), 7Z (2) Allocation bias (allocation concealment), 7Z (3) Detection bias (blinding), 8X Congruency data and methods, 8Z (1) Attrition bias (incomplete outcome data), 8Z (2) Reporting bias (Selective outcome reporting), 9X Presentation of limitations, 10X Conflict of interest, 10Z Publication bias (influence)

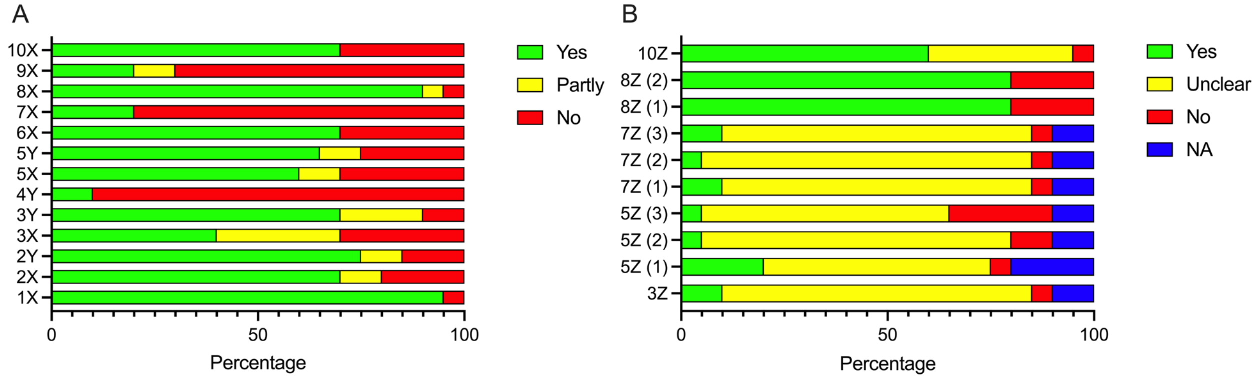


Fig. 2 Presentation example of many included studies – Overview of results. Fictitious data. 1X Peer review, 2X Bench-top, 2Y Bench-top, 3X Animals, 3Y Animals, 3Z Selection bias (baseline characteristics), 4Y Sample size, 5X in vivo design and performance, 5Y in vivo design and performance, 5Z (1) Selection bias (Sequence generation), 5Z (2) Performance bias (Random housing), 5Z (3) Detection bias (Random outcome assessment), 6X Animal welfare compliance, 7X Blinding, 7Z (1) Performance bias (Blinding), 7Z (2) Allocation bias (allocation concealment), 7Z (3) Detection bias (blinding), 8X Congruency data and methods, 8Z (1) Attrition bias (incomplete outcome data), 8Z (2) Reporting bias (Selective outcome reporting), 9X Presentation of limitations, 10X Conflict of interest, 10Z Publication bias (influence)

results was based on both McHugh [60] and Landis and Koch [61]. In general, we found high interrater agreement, ranging from 84.5% to 100%, with a mean of 92.9% (95% CI 91.0–94.8). The kappa indices ranged between 0.50 and 0.96, with a mean of 0.85 (95% CI 0.79–0.91). The quality of reporting had a mean kappa index of 0.86 (95%-CI 0.78–0.94), the methodological quality had a mean kappa index of 0.83 (95%-CI 0.71–0.93) and SYRCLE's risk of bias items had a mean kappa index of 0.68 (95%-CI 0.57–0.79) (Table 3).

Interpretation

The most important aspect of CRIME-Q is that it provides transparency over a wide range of literature. CRIME-Q displays the literature within a given field on multiple levels. For instance, if the quality of reporting is low, there is no way to properly assess methodological (technical) quality or risk of bias. CRIME-Q illuminates inconsistencies and weak points in the literature for readers of a given systematic review on preclinical models. Given the individual scoring, it also highlights the strengths and weaknesses of individual studies,

Table 3 Validation of the CRIME-Q was performed based on a systematic review of 114 records on meningioma animal models [62] through agreement percentage and nominal Kappa index

Items	Type	Agreement %	Kappa - index (SE)	McHugh(60) Interpretation	Landis et al.(61) interpretation
1X Peer review	QoR	100%	*	Almost Perfect	Almost Perfect
2X Bench-top	QoR	91.4%	0.85 (0.077)	Strong	Almost Perfect
2Y Bench-top	MQ	94.0%	0.83 (0.081)	Strong	Almost Perfect
3X Animals	QoR	87.9%	0.84 (0.075)	Strong	Almost Perfect
3Y Animals	MQ	91.4%	0.78 (0.080)	Moderate	Substantial
3Z Selection bias (baseline characteristics)	RoB	94.0%	0.77 (0.073)	Moderate	Substantial
4Y Sample size	MQ	99.1%	*	Almost perfect	Almost Perfect
5X *in vivo* design and performance	QoR	91.4%	0.92 (0.081)	Almost perfect	Almost Perfect
5Y *in vivo* design and performance	MQ	95.3%	0.87 (0.077)	Strong	Almost Perfect
5Z (1) Selection bias (Sequence generation)	RoB	82.8%	0.62 (0.062)	Moderate	Substantial
5Z (2) Performance bias (Random housing)	RoB	93.1%	0.57 (0.076)	Weak	Moderate
5Z (3) Detection bias (Random outcome assessment)	RoB	86.2%	0.71 (0.076)	Moderate	Substantial
6X Animal welfare compliance	QoR	97.4%	0.96 (0.093)	Almost perfect	Almost perfect
7X Blinding	QoR	97.4%	0.82 (0.091)	Strong	Almost perfect
7Z (1) Performance bias (blinding)	RoB	87.1%	0.53 (0.063)	Weak	Moderate
7Z (2) Allocation bias (allocation concealment)	RoB	87.9%	0.50 (0.065)	Weak	Moderate
7Z (3) Detection bias (blinding)	RoB	85.3%	0.50 (0.064)	Weak	Moderate
8X Congruency data and methods	QoR	93.1%	0.66 (0.079)	Moderate	Substantial
8Z (1) Attrition bias (incomplete outcome data)	RoB	96.6%	0.90 (0.081)	Strong	Almost Perfect
8Z (2) Reporting bias (Selective outcome reporting)	RoB	93.1%	0.80 (0.086)	Strong	Substantial
9X Presentation of limitations	QoR	93.1%	0.86 (0.081)	Strong	Almost perfect
10X Conflict of interest	QoR	97.4%	0.95 (0.090)	Almost perfect	Almost perfect
10Z Publication bias (influence)	RoB	94.8%	0.90 (0.089)	Strong	Almost perfect

SE standard error, *QoR* Quality of Reporting, *MQ* Methodological Quality, *RoB* Risk of Bias

* Kappa index cannot be calculated due to too high agreement and/or too few rating items. The data reliability scores (interpretation) were based on McHugh(60) and Landis and Koch [61]. The full scoring from both reviewers is available in the published material of [18] online at: https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-023-04620-7#Sec29 (CRIME-Q full spread sheet)

allowing readers and users of the systematic review to know which studies to be wary of and in which aspects. This enables them to decide for themselves how to interpret the literature.

Discussion

There is a need for a critical appraisal tool that includes quality of reporting, methodological (technical) quality, and risk of bias assessment for all animal studies interventional and noninterventional, which also contains the laboratory work related to the model. Despite a larger focus on proper reporting through various tools, reporting remains lacking in animal research [11, 63]. The CRIME-Q tool unifies various tools and checklists and items of own making in a condensed format. The purpose of the tool is to have a single instrument for critical appraisal with a risk of bias, applicable to all old, new, high-quality, and low-quality preclinical studies on multiple levels. It also serves to provide readers of a systematic review with transparency regarding the included literature.

All the items in CRIME-Q displayed an interrater percentage agreement > 80%, which is considered the minimum acceptable interrater agreement [60]. Cohen's kappa index for the CRIME-Q items showed the strongest interrater reliability for quality of reporting, with 0.86 (95%-CI 0.78–0.94). This did not include items 1X and 4X, where a nearly 100% agreement gave an insufficient number of rating categories to allow Cohen's kappa index to be calculated. CRIME-Q methodological quality showed a similar kappa index of 0.83 (95% CI 0.71–0.93). Regarding interpretation, McHugh [60] suggests a stricter interpretation than Cohen's[61], but using this approach, we still found high interrater reliability ranging from moderate (kappa 0.66) to almost perfect (kappa 0.96) for the quality of reporting and methodological quality items analyzed. Finally, SYRCLE's risk of bias items showed the lowest interrater agreement percentage, despite being quite high of 90.7% (95%-CI 87.5–94.0). The kappa index of the risk of bias items were the lowest 0.68 (95%-CI 0.57–0.79), especially the blinding items 7Z (1–3) which ranged from 0.50–0.53 and 5Z (2) performance bias (random housing). This shows the risk of bias items are more difficult to assess for our data set on meningioma animal models, and require more discussion afterwards before final verdict.

The predetermined list/definition of information to be included for proper assessment of the items results in high interrater agreement, especially for quality of reporting. The quality of reporting is crucial in assessing replicability and transparency, which are often poor in animal research [11]. It is important to acknowledge, there have been minor improvements, especially with better descriptions of randomization, blinded experimental conduct, blinding assessment of outcome, and sample size calculation, but still not sufficient [11]. This emphasizes the need to focus on designing, conducting, and reporting animal studies. Methodological quality is based on the reviewer/assessor group's definition of a well-performed study's technical aspects, and as previously described, a predetermined definition of quality for each MQ item is needed to create transparency. Readers should be able to clearly understand why a certain score was assigned and based on which information, as outlined in the predetermined list. Here, they should also be able to evaluate how the assessor group defined their criteria and decide whether they agree with it or not. The lower Cohen's kappa indices of Y (MQ) compared to those of their X (QoR) counterparts emphasize the need for a solid predetermined definition of quality for the reviewers to follow when assessing. A suboptimal definition will mean a greater chance of disagreements and a greater need for discussions afterwards. Methodological quality items help assess critical items such as laboratory work for the animal model, choice of animals, and in vivo design and performance. Methodological quality differs from risk of bias [15] in the CRIME-Q. While risk of bias focuses on study design and validity - the likelihood that features of the study design or conduct of the study will yield misleading results - methodological quality focuses on technical performance. Hence, CRIME-Q assesses multiple aspects of the studies.

SYRCLE's RoB is based on the Cochrane Collaboration's tool for assessing the risk of bias in clinical randomized control trials [17] and is intended to be used in interventional animal studies. Most of the items from SYRCLE's RoB tool are primarily rated 'Unclear' in our own systematic review on meningioma animal models [18]. We believe this is likely because of the quality of reporting issues in the literature [64], but also study design flaws[65], which makes SYRCLE's risk of bias unsuitable for assessing overall quality as a standalone method/tool given the low level of reporting and lack of published/accessible in vivo protocols. However, with CRIME-Q, the quality of reporting and methodological quality can still be addressed in these studies for a nuanced transparent assessment of the literature—or, at the very least, identifying the specific step where the literature falls short: in reporting quality, methodological quality, or both. Furthermore, some noninterventional studies cannot be adequately assessed using SYRCLE's RoB alone due to the tool's design and intended purpose, which often necessitates recording certain items as 'not applicable.' Items such as allocation to groups, random housing, random selection for outcome assessment, and blinding may not apply in many cases. However, in these cases SYRCLE's RoB still addresses key aspects of study validity, such as attrition (incomplete outcome data) and reporting bias (selective outcome reporting), ensuring a more comprehensive evaluation. We included SYRCLE's RoB in CRIME-Q because it is an excellent validated tool. All the items included should be applied whenever possible. It is important to consider instances where biases could be applied—even, especially biases we cannot control, which may affect outcomes—e.g., blinding in surgical studies. Surgeons cannot be blinded but may still present a risk of bias due to lack of blinding.

It can be difficult to judge the overall quality of a study/record, and there have been many suggestions as to which critical appraisal tools to use [13]. However, it is a crucial step in performing systematic reviews [12], which could be the missing link in translational animal science and help propel it forward [5]. Most available tools are not validated and are specified to certain situations and disease models such as Minnerup et al. [23] and Sena et al. [30], which are used for stroke models, Rice et al. [28] for pain, Marshall et al. [66], for sepsis, who offer a short checklist without much explanation. Furthermore, there are tools intended to be used in preclinical drug research primarily [32, 36], which lessens generalizability. CRIME-Q was developed with inspiration from some of the validated and recommended tools for reporting and assessing the overall quality of studies to ensure uniformity in future research [10, 14–16]. The tool is designed for assessing the overall quality of studies, but it has the potential to be applied to the development, design, and performance of animal studies to ensure that all aspects are covered before study initiation and publication. ARRIVE 2.0 remains the recommended tool for reporting animal studies, but it does not cover the laboratory work for the model; elements from CRIME-Q could thus supplement ARRIVE 2.0. A limitation of CRIME-Q is that it has not yet been externally validated, and we urge other groups to test the tool and provide us with feedback through the corresponding author information listed. Moreover, there are no scoring systems available for the CRIME-Q since there is no clear way to label and weigh the various items; e.g., peer review, despite being standard in research, is not always indicative of high quality. The tool is meant to present included literature in a nuanced (QoR, MQ and RoB) and transparent manner. Finally, similar to other tools, including SYRCLE's RoB [15, 67, 68], we hope and expect user feedback to allow us to update and adapt the tool accordingly.

Conclusion

The CRIME-Q tool unifies multiple aspects of the quality of reporting, methodological (technical) quality, and SYRCLE's risk of bias from the laboratory to in vivo design and performance in animal studies. It contains both items inspired by older methods and contains unique items of own making. The tool allows overall assessment of studies included in systematic reviews on animal research with high interrater agreement percentage and moderate-almost perfect interrater reliability through Cohen's kappa through internal validation. The CRIME-Q tool is applicable to both older research that lacks proper reporting and newer studies and is not restricted by model or study design type (interventional–noninterventional). It can also be useful when designing and conducting animal experiments to ensure proper reporting and design in terms of replicability, transparency, and validity. We acknowledge that CRIME-Q has not yet been externally validated, and we urge other groups to test the tool.

Abbreviations

CRIME-Q	Critical Appraisal of Methodological (technical) Quality, Quality of Reporting and Risk of Bias in Animal Research (CRIME-Q)
QoR	Quality of reporting
MQ	Methodological quality
RoB	Risk of bias
3R	Reduction, refinement, replacement

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-024-02413-0.

Supplementary Material 1.

Acknowledgements

We would like to thank Prof. Asbjørn Hróbjartsson, Head of the Cochrane Denmark & Centre for Evidence-Based Medicine Odense (CEBMO) for statistical assistance and Claire Gudex, University of Southern Denmark, for final proofreading.

Authors' contributions

MSA: first draft, conceptualization, methodology, figure, and tables, review, writing, editing, and approved the final draft; MSK: conceptualization, methodology, review, writing, editing, and approved the final draft; ASP-M: conceptualization, methodology, review, writing, editing, and approved the final draft; FRP: conceptualization, methodology, review, writing, editing, and approved the final draft; BH: methodology and items regarding in vivo items, review, writing, editing, and approved the final draft; BBO; methodology and items regarding in vitro items, review, writing, editing, and approved the final draft. CBP: review, writing, editing, and approved the final draft, TM: review, writing, editing, and approved the final draft, CM: review, writing, editing, and approved the final draft

Funding

No funding related to the current study was received.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Neurosurgery, Odense University Hospital, Odense, Denmark. [2]Department of Clinical Research, University of Southern Denmark, Odense, Denmark. [3]BRIDGE (Brain Research - Inter Disciplinary Guided Excellence), University of Southern Denmark, Odense, Denmark. [4]Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark. [5]Centre for Evidence-Based Medicine Odense (CEBMO), Copenhagen, Denmark. [6]NHTA: Market Access & Health Economics Consultancy, Copenhagen, Denmark. [7]Department of Neurosurgery, Rigshospitalet, Copenhagen, Denmark. [8]Copenhagen University, Copenhagen, Denmark. [9]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [10]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [11]Department of Neuropathology, Otto-Von-Guericke University, Magdeburg, Germany. [12]Department of Nuclear Medicine, Odense University Hospital, Odense, Denmark. [13]Department of Surgical Pathology, Zealand University Hospital, Roskilde, Denmark.

Received: 10 July 2023 Accepted: 11 November 2024
Published online: 19 December 2024

References

1. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg. 2011;128(1):305–10.
2. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol. 2018;18(1):5.
3. 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(3):427–37.
4. Leenaars M, Hooijmans CR, van Veggel N, ter Riet G, Leeflang M, Hooft L, et al. A step-by-step guide to systematically identify all relevant animal studies. Lab Anim. 2012;46(1):24–31.
5. 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(3):e89981.
6. Hooijmans CR, Ritskes-Hoitinga M. Progress in using systematic reviews of animal studies to improve translational research. PLoS Med. 2013;10(7):e1001482.
7. Zhu H, Jia Y, Leung SW. Citations of microRNA biomarker articles that were retracted: a systematic review. JAMA Netw Open. 2024;7(3):e243173.
8. Van Noorden R. How big is science's fake-paper problem? Nature. 2023;623(7987):466–7.
9. Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke. 2008;39(10):2824–9.
10. du Percie Sert N, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, et al. Reporting animal research: explanation and elaboration for the ARRIVE guidelines 2.0. PLoS Biol. 2020;18(7):e3000411.
11. Kousholt BS, Praestegaard KF, Stone JC, Thomsen AF, Johansen TT, Ritskes-Hoitinga M, et al. Reporting quality in preclinical animal experimental research in 2009 and 2018: a nationwide systematic investigation. PLoS One. 2022;17(11):e0275962.
12. Bahadoran Z, Mirmiran P, Kashfi K, Ghasemi A. Importance of systematic reviews and meta-analyses of animal studies: challenges for animal-to-human translation. J Am Assoc Lab Anim Sci. 2020;59(5):469–77.

13. Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environ Health Perspect. 2013;121(9):985–92.
14. Cramond F, Irvine C, Liao J, Howells D, Sena E, Currie G, et al. Protocol for a retrospective, controlled cohort study of the impact of a change in Nature journals' editorial policy for life sciences research on the completeness of reporting study design and execution. Scientometrics. 2016;108:315–28.
15. 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.
16. Macleod MR, O'Collins T, Howells DW, Donnan GA. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke. 2004;35(5):1203–8.
17. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
18. Andersen MS, Kofoed MS, Paludan-Muller AS, Pedersen CB, Mathiesen T, Mawrin C, et al. Meningioma animal models: a systematic review and meta-analysis. J Transl Med. 2023;21(1):764.
19. Vesterinen HM, Egan K, Deister A, Schlattmann P, Macleod MR, Dirnagl U. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the Journal of Cerebral Blood Flow and Metabolism. J Cereb Blood Flow Metab. 2011;31(4):1064–72.
20. Conrad JW Jr, Becker RA. Enhancing credibility of chemical safety studies: emerging consensus on key assessment criteria. Environ Health Perspect. 2011;119(6):757–64.
21. Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. Improving the translational hit of experimental treatments in multiple sclerosis. Mult Scler. 2010;16(9):1044–55.
22. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. J Pharmacol Pharmacother. 2010;1(2):94–9.
23. Minnerup J, Wersching H, Diederich K, Schilling M, Ringelstein EB, Wellmann J, et al. Methodological quality of preclinical stroke studies is not required for publication in high-impact journals. J Cereb Blood Flow Metab. 2010;30(9):1619–24.
24. Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. Altern Lab Anim. 2010;38(2):167–82.
25. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? PLoS Med. 2010;7(3):e1000245.
26. Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, et al. Good laboratory practice: preventing introduction of bias at the bench. Stroke. 2009;40(3):e50–2.
27. Fisher M, Feuerstein G, Howells DW, Hurn PD, Kent TA, Savitz SI, et al. Update of the stroke therapy academic industry roundtable preclinical recommendations. Stroke. 2009;40(6):2244–50.
28. Rice ASC, Cimino-Brown D, Eisenach JC, Kontinen VK, Lacroix-Fralish ML, Machin I, et al. Animal models and the prediction of efficacy in clinical trials of analgesic drugs: a critical appraisal and call for uniform reporting standards. Pain. 2008;139(2):243–7.
29. Sniekers YH, Weinans H, Bierma-Zeinstra SM, van Leeuwen JP, van Osch GJ. Animal models for osteoarthritis: the effect of ovariectomy and estrogen treatment - a systematic approach. Osteoarthritis Cartilage. 2008;16(5):533–41.
30. Sena E, van der Worp HB, Howells D, Macleod M. How can we improve the pre-clinical development of drugs for stroke? Trends Neurosci. 2007;30(9):433–9.
31. Hobbs DA, Warne MS, Markich SJ. Evaluation of criteria used to assess the quality of aquatic toxicity data. Integr Environ Assess Manag. 2005;1(3):174–80.
32. van der Worp HB, de Haan P, Morrema E, Kalkman CJ. Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. J Neurol. 2005;252(9):1108–14.
33. de Aguilar-Nascimento JE. Fundamental steps in experimental design for animal studies. Acta Cir Bras. 2005;20(1):2–8.
34. Bebarta V, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med. 2003;10(6):684–7.
35. Verhagen H, Aruoma OI, van Delft JH, Dragsted LO, Ferguson LR, Knasmuller S, et al. The 10 basic requirements for a scientific paper reporting antioxidant, antimutagenic or anticarcinogenic potential of test substances in in vitro experiments and animal studies in vivo. Food Chem Toxicol. 2003;41(5):603–10.
36. Lucas C, Criens-Poublon LJ, Cockrell CT, de Haan RJ. Wound healing in cell studies and animal model experiments by Low Level Laser Therapy; were clinical studies justified? a systematic review. Lasers Med Sci. 2002;17(2):110–34.
37. Festing MF, Altman DG. Guidelines for the design and statistical analysis of experiments using laboratory animals. ILAR J. 2002;43(4):244–58.
38. Johnson PD, Besselsen DG. Practical aspects of experimental design in animal research. ILAR J. 2002;43(4):202–6.
39. Horn J, de Haan RJ, Vermeulen M, Luiten PG, Limburg M. Nimodipine in animal model experiments of focal cerebral ischemia: a systematic review. Stroke. 2001;32(10):2433–8.
40. Klimisch HJ, Andreae M, Tillmann U. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol. 1997;25(1):1–5.
41. Hsu DW, Efird JT, Hedley-Whyte ET. Progesterone and estrogen receptors in meningiomas: prognostic considerations. J Neurosurg. 1997;86(1):113–20.
42. Agerstrand M, Kuster A, Bachmann J, Breitholtz M, Ebert I, Rechenberg B, et al. Reporting and evaluation criteria as means towards a transparent use of ecotoxicity data for environmental risk assessment of pharmaceuticals. Environ Pollut. 2011;159(10):2487–92.
43. National-Research-Council-(US)-Institute-for-Laboratory-Animal-Research. Guidance for the Description of Animal Research in Scientific Publications. The National Academies Collection: Reports funded by National Institutes of Health: Washington (DC); 2011.
44. Lamontagne F, Briel M, Duffett M, Fox-Robichaud A, Cook DJ, Guyatt G, et al. Systematic review of reviews including animal studies addressing therapeutic interventions for sepsis. Crit Care Med. 2010;38(12):2401–8.
45. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measur. 1960;20:37–46.
46. Gannot G, Cutting MA, Fischer DJ, Hsu LJ. Reproducibility and transparency in biomedical sciences. Oral Dis. 2017;23(7):813–6.
47. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester (UK): John Wiley & Sons; 2019.
48. Kelly J, Sadeghieh T, Adeli K. Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. EJIFCC. 2014;25(3):227–43.
49. Smith R. Peer review: a flawed process at the heart of science and journals. J R Soc Med. 2006;99(4):178–82.
50. Tennant JP, Ross-Hellauer T. The limitations to our understanding of peer review. Res Integr Peer Rev. 2020;5:6.
51. Charan J, Kantharia ND. How to calculate sample size in animal studies? J Pharmacol Pharmacother. 2013;4(4):303–6.
52. Fitts DA. Ethics and animal numbers: informal analyses, uncertain sample sizes, inefficient replications, and type I errors. J Am Assoc Lab Anim Sci. 2011;50(4):445–53.
53. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012;23(5):524–32.
54. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS OnE. 2008;3(8):e3081.
55. Gauthier C, Griffin G. Using animals in research, testing and teaching. Rev Sci Tech. 2005;24(2):735–45.
56. Karp NA, Pearl EJ, Stringer EJ, Barkus C, Ulrichsen JC, du Percie Sert N. A qualitative study of the barriers to using blinding in in vivo experiments and suggestions for improvement. PLoS Biol. 2022;20(11):e3001873.
57. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci U S A. 2018;115(11):2600–6.
58. Ross PT, Bibler Zaidi NL. Limited by our limitations. Perspect Med Educ. 2019;8(4):261–4.

59. Vineis P, Saracci R. Conflicts of interest matter and awareness is needed. J Epidemiol Community Health. 2015;69(10):1018–20.
60. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–82.
61. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74.
62. Andersen MS, Kofoed MS, Paludan-Müller AS, Pedersen CB, Mathiesen T, Mawrin C, et al. Meningioma animal models: a systematic review and meta-analysis. J Transl Med. 2023;Accepted.
63. Chitnis KR, Shah AC, Jalgaonkar SV. A Study to assess the quality of reporting of animal research studies published in pubmed indexed journals: a retrospective, cross-sectional content analysis. Cureus. 2022;14(1):e21439.
64. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS One. 2009;4(11):e7824.
65. Bailoo JD, Reichlin TS, Wurbel H. Refinement of experimental design and conduct in laboratory animal research. ILAR J. 2014;55(3):383–91.
66. Marshall JC, Deitch E, Moldawer LL, Opal S, Redl H, van der Poll T. Preclinical models of shock and sepsis: what can they tell us? Shock. 2005;24(Suppl 1):1–6.
67. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.
68. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

Publisher's Note