Systematic Review and Meta-Analysis of Methodological Quality in *In Vivo* Animal Studies of Subarachnoid Hemorrhage

Basil E. Grüter, MD (0000-0002-6314-2482)[1,2] Davide Croci, MD (0000-0002-7436-4267)[2] Salome Schöpf, MD[2] Edin Nevzati, MD (0000-0001-8943-0290)[2] Donato d'Allonzo, MD (0000-0002-6759-2106)[1], Jacqueline Lattmann, BSc[1] Tabitha Roth, BSc[1] Benjamin Bircher, BSc[1] , Carl Muroi, MD (0000-0001-9999-3345)[2] , Gilles Dutilh, PhD (0000-0002-6719-2508)[3] Hans Rudolf Widmer, PhD (0000-0003-3378-8765)[4] Nikolaus Plesnila, MD (0000-0001-8832-228X)[5] Javier Fandino, MD (0000-0002-5375-3087)[1,2] Serge Marbacher, MD, PhD (0000-0001-6305-7571)[1,2]

[1] Department of Neurosurgery, Kantonsspital Aarau, Switzerland

[2] Cerebrovascular Research Group, Department for BioMedical Research, University of Bern, Bern, Switzerland

[3] Department of Clinical Research, Clinical Trial Unit, University of Basel Hospital, Basel, Switzerland.

[4] Department of Neurosurgery, University Hospital Bern, Bern, Switzerland

[5] Institute for Stroke and Dementia Research (ISD), Ludwig-Maximilians University Munich, Germany

Corresponding author: Basil E. Grüter, MD, Department of Neurosurgery c/o Neuro Research Office, Kantonsspital Aarau, Tellstrasse 1, 5001 Aarau, Switzerland
+41 62 838 66 34. Email: basil.grueter@ksa.ch, neurosurgery@ksa.ch

Acknowledgements: We thank Lea Vogt for providing hard copies of more than 1000 articles eligible for full text analysis. We express our special thanks to Sarah McCann (CAMARADES Methodologist, Centre for Clinical Brain Sciences, University of Edinburgh, UK) who supported us in data management and data extraction. We are also thankful to Mary Kempner from Glia Media for medical editing.

Abstract

As a result of increased awareness of wide-spread methodological bias and obvious translational roadblocks in subarachnoid haemorrhage (SAH) research, various checklists and guidelines were developed over the past decades. This systematic review assesses the overall methodological quality of preclinical SAH research. An electronic search for pre-clinical studies on SAH revealed 3415 potential articles. Of these, 765 original research papers conducted *in-vivo* in mice, rats, rabbits, cats, dogs, pigs, goats and non-human primates with a focus on brain damage related to delayed cerebral vasospasm and early brain injury met the inclusion criteria. We found methodological shortcomings still to prevail in preclinical SAH research. In addition, basic animal characteristics were typically well described but important technical parameters of SAH induction were often underreported. None of the species, models, or techniques used in preclinical SAH research was methodologically superior to the others. Methodological quality of preclinical SAH research was independent of the number of citations or impact factor of a publication. Consequently, we suggest the SAH research community should consider strategies to improve pre-clinical research quality in their field, such as public platforms to (pre) register preclinical experiments, consequent support of open science policies, stricter editorial (and reviewer) control of (pre)existing guidelines and increased efforts in education and training of good laboratory practice for the next generation of researchers.

Key Words: Animal model; Delayed cerebral vasospasm; Early brain injury; In-vivo preclinical research; Methodological quality; Subarachnoid hemorrhage.

Introduction

Preclinical animal experiments have contributed much toward our understanding of the pathophysiology of subarachnoid hemorrhage (SAH) and the development of treatment concepts for early brain injury (EBI) and delayed cerebral vasospasm (DCVS) in humans. Importantly, beneficial effects achieved in animal models commonly do not translate into improved functional outcomes in patients.[1-3] The various obstacles responsible for the translational roadblocks can arise throughout the course of scientific experimentation, that is, from planning to design, execution, and reporting.[4, 5] Therefore, results not only relate to the tested variable but to the study design and quality. For instance, a systematic review on pharmacologic treatments that reduced vasospasm in animal models of SAH reported publication bias that ultimately resulted in an overestimation of the true effect of these drugs.[1] While deliberate fraud has occurred on rare occasions (<1%), selective reporting of results (e.g., omission of data) is more common and may often explain the underlying poor reproducibility of animal studies.[6] Furthermore, previous investigations on ischemic stroke found a negative correlation between described effect size in preclinical studies and methodological quality.[5, 7-9] Specifically, as methodological quality of an investigation increases, the studied effect decreases and vice versa: in fact, poorly conducted studies may more likely spur the clinical evaluation of new treatment strategies. A lack of standardization of SAH models [10, 11] and poor methodological quality of experimental studies [6, 12] constitute substantial problems in preclinical research.

To avoid systematic bias and to increase reproducibility of preclinical stroke trials, various platforms, checklists, and guidelines have been developed.[13] Among them is a 10-item checklist from the UK-based CAMARADES platform (Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies). Released in 2007, the CAMARADES platform was derived in part from pre-existing checklists, from aspects known to be important for study quality in clinical trials and from knowledge of the potential synergistic effects of unintended hypothermia with anesthetic used.[14] In 2008, Macleod published a landmark instruction to “good laboratory practice” in preclinical stroke research.[15] Since the publication of the ARRIVE guidelines (Animals in Research: Reporting In Vivo Experiments) in 2010, this guideline became the cornerstone for reporting animal research [16] and has been promoted by more than 1000 journals as a prerequisite standard for publication.[17]

Despite these efforts, adherence to quality standards in methodology for preclinical *in vivo* stroke research remains a disputed topic among scientists and clinicians.[4, 13, 18, 19] Therefore, this meta-analysis aims to portray the overall methodological quality of preclinical

SAH literature, as well as reported descriptions of basic animal characteristics and relevant parameters of SAH-induction techniques among various species and SAH models. Potential associations of methodological quality are evaluated related to impact factor and number of citations of published studies.

Material and Methods

Literature search

A PubMed literature research was performed using a previously reported search strategy.[10] Briefly, we included all studies published between January 1, 2000 and January 31, 2015 using the key words “murine”, “rat”, “rabbit”, “canine”, “primate”, “cat”, “pig”, and “goat” in combination with “subarachnoid hemorrhage”. The search was automatically restricted to animals by using the MEDLINE PubMed limit “animals”. Two investigators (SM and CM) independently screened titles and abstracts for eligible studies and removed duplicates. Any discrepancy in selecting a study between these two authors was decided by a third author, and if necessary, the full text was read to determine its eligibility. Potential studies then underwent full text screening (BB, BEG, DC, EN, JL, SS, TR) to confirm inclusion to the data system of CAMARADES. Uncertainties in full text extraction and assessment of methodological quality characteristics were discussed with a third author (SM). The search algorithm followed PRISMA guidelines (Figure 1). [10, 20]

Eligibility criteria

We considered all preclinical SAH studies conducted *in vivo* in mice, rats, rabbits, cats, dogs, pigs, goats and non-human primates with a study aim toward intracranial consequences of SAH (e.g., EBI or DCVS). Excluded were studies on extracranial vessels or organs other than the brain (e.g., heart or lung), studies without SAH, studies with agents causing vasoconstriction or brain damage other than whole blood, *in vitro* experiments, methodological studies (e.g., description and comparison of new models or refinements on pre-existing models), and review articles. Furthermore, non-original research, conference papers, and research articles written in a language other than English were also excluded from analysis.

Analyzed features

The quality of preclinical SAH studies was assessed using a 10-point methodological quality score (MQS) based on the CAMARADES checklist [14] with slight adaptations. Items scored included: (1) sample size calculation, (2) random allocation of animals to treatment group, (3) blinded induction of injury, (4) monitoring of physiological variables, (5) control of temperature, (6) mortality reported, (7) blinded assessment of outcome, (8) statement of

potential conflicts of interest, (9) statement of compliance with animal welfare regulations, and (10) publication in a peer-reviewed journal. Inclusion of any item received 1 point (maximum 10, range 0-10). The following methodological practices important for preclinical SAH research were also assessed: use of neuroprotective anesthetics (i.e., defined as barbiturates and benzodiazepines because they increase tolerance of neurological tissues to ischemia and change intra-cellular responses to energy supply deprivation); use of comorbid animals; animals fasted before surgery; blood gas analysis; reasons for excluding animals from final analysis; and reports of morbidity. Basic animal characteristics reported were defined by strain, sex, age, and weight of experimental animals. Finally, technical parameters of SAH induction, known to significantly influence outcome,[10] were assessed for injection volume, injection time, filament size, and anesthetics used. Cross-referencing to an earlier publication for methodological characteristics was limited to no more than one link back in order to minimize reporting bias. Thus, if the given reference referred to a further reference, the item was considered not given for the first analyzed paper. For all studies in the analysis, the number of citations as of May 2018 was investigated on Google scholar (Google LLC, California, USA; https://scholar.google.com) and the impact factor was defined by the journal of publication or by Clarivate Analytics (Philadelphia, USA; https://clarivate.com).

Standard models

The above-mentioned analyses performed for all studies were then itemized specifically by animal species and most often performed (standard) SAH induction techniques. Standard techniques, defined as the most frequently used techniques, are extensively discussed elsewhere.[10] Briefly, we resumed the following definitions: Mouse T1: Single injection infra- and supratentorial of 0.06 ml blood during 15 seconds (s); Mouse T3: Endovascular perforation of a supratentorial vessel with a 5-0 suture; Rat T1 single: Single injection infra- and supratentorial of 0.3 ml (0.1 ml/kg) blood during 20 s; Rat T1 double: Infratentorial injection of 0.3 ml (0.1 ml/kg) blood during 120 s with repetition of the same procedure after 48 hours; Rat T3: Endovascular vessel perforation supratentorial with a 4-0 suture; Rabbit T1 single: Infratentorial injection of 1 ml (1ml/kg) blood during 60 s; Rabbit T1 double: Infratentorial injection of 1.5 ml (1.5ml/kg) during 60 s and repetition of the same procedure after 48 hours; Dog: Infratentorial injection of 0.5 ml (0.5ml/kg) during 60 s and repetition of the same procedure after 48 hours; primate: Craniotomy and supratentorial clot placement with 5 ml blood.[10]

Data collection and analysis

The entire study protocol was preregistered on the CAMARADES platform and has been publicly accessible since September 1st 2014 on

http://www.dcn.ed.ac.uk/camarades/research.html#protocols. All data were extracted to the CAMARADES Microsoft Access 2003 data manager (Windows, Microsoft, Seattle, USA) and analyzed using R version 3.4.3. In our descriptive analyses, data are summarized in tables and figures. For comparisons of methodological quality between studies using standard models and other models, Wilcoxon signed rank tests are reported. Associations between methodological quality score (MQS), impact factor, and number of citations were assessed by Spearman correlations and regression analysis. A p-value <0.05 was considered significant.

Results

Of the 3415 potential articles identified in our electronic search for SAH animal studies, 765 studies were included for meta-analysis. Excluded from final analysis were 1583 studies after title and abstract screening, 997 after full text reviews, and 70 studies during data extraction (Figure 1).

For methodological quality score (MQS), the median value was 4 (range 1-7) for studies published in 2000 and 5 (range 1-9) in 2015, with a discontinuous peak after publication of ARRIVE guidelines [16] in 2010 (Figure 2). Adherence to the individual methodological study properties that form the MQS varied widely. Of all 765 studies, 761 (99.5%) were published in a peer-reviewed journal and 660 (86.3%) studies confirmed compliance with animal welfare regulations. However, only 0.9% (6 rat studies and 1 mouse study) defined a sample size calculation. Reports of a potential conflict of interest steadily increased from none in 2000 to 51.2% (41/80) in 2011 (Figure 3). With publication of ARRIVE guidelines in 2010, there was increased reporting of random allocation to groups, blinded induction of injury, mortality, and more recently, a decrease in blinded assessment of outcome.

Other methodological practices not reflected in the MQS were equally and poorly reported among 765 studies. These included morbidity noted in only 72 studies (9.5%), reasons for animal exclusion from final analysis in 78 studies (10.2%), use of comorbid animals in 23 studies (3%), avoidance of neuroprotective anesthetics in 191 studies (25%), and indication of fasting before surgery in 36 studies (4.7%). Blood gas analysis was more often performed in dogs (63/101, 62.4%), pigs (3/4, 75%), and primates (15/25, 60%) and less often in mice (10/56, 17.9%) and rats (167/411, 40.6%) (Supplementary Figure 1). Methodological quality did not significantly vary among experiments by species (Figure 4) and was not associated with geographic origin of the researchers (Supplementary Figure 2). There was no association between MQS and specific SAH induction techniques (Supplementary Figure 3).

In sub-analysis of the various types of SAH induction techniques by species, MQS was significantly higher for studies using rats in a standard rather than non-standard model (p=0.018) (Supplementary Figures 4-8). See Table 1 for relative frequencies of MQS parameters for all studies and itemized by species, and Supplementary Table 1 for itemization per standard SAH model specifically for each species.

Although reporting was insufficient for methodological quality in preclinical SAH studies, animal characteristics of paramount importance for study outcome [21], such as strain, sex, age, weight, and narcotics use, were more often specified. Of the 765 included studies, 665 (87%) reported strain, 595 (78%) reported sex, 718 (94%) reported age and/or weight of the experimental animals, and 687 (90%) reported narcotics used (Table 2). No relevant differences in study quality were found between different protocols for anesthetics used (Figure 5). Although 707 (92%) studies described the application of either an injection (534 studies) or filament (173 studies) SAH-induced model [10], parameters of the induction techniques were underreported. Specifically, only 235 (33%) reported the volume of injected blood. Likewise, only 44% (235 of 534 studies) using an injection model stated blood injection times and 43% (75 of 173 studies) using a filament model stated filament size used for SAH induction (Supplementary Table 2).

Journal impact factor for all studies averaged 3.3 (range 0.0-38.2). Better adherence to methodological quality markers, as reflected by a higher MQS, was not associated with publication in a higher impact factor (supposedly higher quality) journal (p=0.174) (Figure 6). Higher MQS of a study was also not associated with higher numbers of citations (p=0.724) (Figure 7). In an additional regression analysis testing whether MQS might affect the number of citations only in higher impact journals, we determined the significance of the interaction between MQS and impact factor as a predictor of number of citations. This interaction factor turned out non-significant. However, a weak (ρ=0.126) but significant (p<0.001) correlation could be confirmed between higher impact factor and number of citations (Supplementary Figure 9).

Discussion

This systematic review and meta-analysis of pre-clinical SAH research in 765 publications showed overall poor methodological quality with only moderate improvement over the 15-year study period (2000-2015). Technical considerations of SAH-induction techniques, which are paramount for interpretation of outcome of preclinical studies, were also underreported. Poor methodological quality was independent of animal species and SAH model. Lastly, better adherence to methodological quality was neither associated with higher impact factor nor more frequent citation of these studies.

Possible reasons for the poor methodological quality reporting

Poor methodological quality is a problem of preclinical research in general and not specific for the field of SAH research.[22] For instance, reproducibility of preclinical studies that identified new drug targets ranged from 0% to 32%.[23-25] Specifically, methodological shortcomings in experimental studies have been criticized in meta-analyses of various subtypes of ischemic stroke [9, 26, 27] and intracerebral hemorrhage.[4] However, this low methodological quality must be interpreted cautiously. For example, a lack of description about experimental planning or setting did not necessarily mean they were not performed but were possibly excluded to adhere to the strict word counts specified by the journal.[4] Low impact factor journals may not allow any supplementary materials (at least for methods) and if not explicitly demanded by editors or reviewers, methodological details may sometimes be sacrificed in favor of conciseness. In another example, the frequency of a conflict of interest statement (COI) increased from literally 0 in 2004 to >50% of all studies in 2011. This change likely arose as more high-quality journals required a COI statement for submission and publication.

Most reviews and meta-analyses on preclinical stroke trials uniformly emphasize the importance of a sample size calculation, including at least the expected difference between groups, expected variance, and desired statistical power.[4, 5, 15, 16] Despite this apparent accordance, very few studies effectively included these calculations. One ethical principle is to use the minimum number of animals required to precisely demonstrate the outcome of interest. As researchers may favor small sample sizes for administrative, timely, and pecuniary considerations, many new animal studies remain underpowered. One option to improve performing and reporting of sample size calculation could be to require such information on submission (e.g., check box). However, most researchers might consider the requirement an additional administrative/regulatory burden and ignore it. Rather than a helpful tool, a recent survey of 302 researchers experienced with multiple years of animal research revealed that >50% who published their latest article in a journal that endorsed the ARRIVE guidelines were unaware of these guidelines.[22] Nevertheless, our meta-analysis revealed a temporary boost of increased reporting of allocation concealment, blinded injury induction, and mortality in the first years after publication of ARRIVE guidelines (prompted by their adoption as required reporting guidelines in most leading journals). Unfortunately, this short-lasting practice declined again to levels of several years earlier. Our finding aligns with a study investigating the implementation of ARRIVE guidelines 2 years after their publication that found only moderate improvement in reporting standards in high-impact studies.[28] Furthermore, and certainly against their intention, other factors not explicitly mentioned in the ARRIVE guidelines (e.g., blinded assessment of outcome) were even less often reported in the years after their publication.

Baseline animal characteristics and methodology on SAH induction

Similar to methodological integrity, reporting of details of the animal model and SAH induction technique is highly relevant for reproducibility and eventually translation into new treatment concepts.[10] Various studies demonstrated the great influence of animal's baseline characteristics (e.g., age, sex, strain, genetic background) and technical parameters of SAH induction (e.g., blood injection velocity, blood volume, filament size) on the measured outcome.[11, 29-33] Our study showed that basic animal characteristics were reported by most studies (77.8% – 86.9%) but relevant information on SAH induction techniques were often missing (33.2%-44.0%).

Reporting of technical details is of utmost importance as illustrated in this case from the field of ischemic stroke: In a randomized blinded pre-clinical multi-center study, the authors found that even with agreement on a common standard for the stroke model, the results from contributing centers may not be fully reproducible by other contributing centers.[34] Given the abundance of SAH induction techniques, all with potential for adaptations and individual refinements, one can only guess how SAH and physiological responses must fundamentally differ among studies and how that impacts the measured study endpoints. After the paradigm shift from DCVS-oriented research towards EBI endpoints, many labs may have continued to conduct their experiments in their established models, some of which were potentially inadequate to investigate this new aspect of SAH.[35, 36]

Interestingly, our meta-analysis did not show relevant differences in the quality of reporting between species. At first glance this may seem surprising. One might expect studies on higher species, such as primates, to be relatively advanced, close to clinical translation, and therefore methodologically strong versus relatively new ideas expected to be methodologically weaker, such as initial pilot studies in mice or rats. Rather, we identified methodologic quality reporting was equally insufficient in both types of studies. Several explanations exist. First, the few studies on higher species that met our inclusion criteria for this meta-analysis may be insufficient for generalizability. Second, studies on higher species became increasingly uncommon during the past 20 years while the number of studies using small rodents increased. In our analysis, most studies on higher species were performed before the launch of guidelines while many studies on smaller species were conducted after publication of these guidelines. Furthermore, our analyses showed that descriptions were comparable for specifying basic animal characteristics and reporting relevant considerations of SAH-induction techniques among the various models. Therefore, we conclude that

attention to methodological study quality is independent from the choice of a specific SAH model and no methodology was superior to another.

Correlation of impact factor with methodological quality

The mean impact factor of 3.3 (range 0.0-38.2) for all included preclinical SAH papers is well below the 4.6 mean (range 0.7-30.6) reported for animal stroke studies on neuroprotective drugs.[37] Data are lacking to compare these values with other subfields of preclinical stroke research. Publication choice and impact factor could be affected by the perception of the disease as specialty. That is, SAH is viewed as a neurosurgical disease unlike ischemic stroke, which is generally associated with the much larger neurological community. Given the target specialty, the trend to publish (preclinical) SAH studies in neurosurgical journals could explain the lower impact factor than that of publishing for a broad audience of a neurological journal. A specialty journal may also face budget or staff restrictions that could influence the adherence to guidelines or provision of supplementary repositories for detailed methodological descriptions; both could affect methodological quality as determined by the MQS.

We concluded that highly ranked articles were, in general, not methodological stronger than articles in lower impact journals. This finding aligns with a systematic review on preclinical stroke trials showing that impact factor was more likely associated with the complexity of the investigation than methodological quality.[37] Even more alarming, our analysis revealed that citation rates of the articles were unrelated to their quality. Such an unfiltered retransmission of methodologic heterogeneously acquired data may partially contribute to the low translation rate of new concepts from bench to bedside. However, a small proportion of the published SAH experiments were also methodologically very strong (MQS 8 and 9), which proves that better methodology is achievable and a realistic goal that authors should aim for.

Study limitations

Our meta-analysis has several limitations. First, with no unquestionable objective measurement of study quality, we chose the pre-defined Macleod criteria [14] and applied an established, approved system. Nonetheless, equal weighting of each factor may be problematic because the MQS summarizes very different entities that affect study quality in varying ways. [9, 26]. Among them, particularly “compliance with animal welfare regulations” and “publication in peer-reviewed journals” are not direct quality markers. Second, certain parameters could not be fully investigated because of lack of resources. For instance, “measurements of physiological variables” was considered positive when mentioned or if a value was defined for heart rate or blood pressure; however, we did not further investigate

other relevant points (e.g., method, time point, frequency, etc.) of measurement that certainly could impact the quality of the study. Third, cross-referencing in describing methodological characteristics of SAH induction was limited to one-tier; therefore, underreporting of these characteristics may be overestimated. For instance, if a study cited a previous publication for detailed technical descriptions of SAH induction, and that study instead cited an earlier publication, we did not follow this second link and the model description was considered negative. Journal length restrictions may also limit presentation of MQS relevant parameters and therefore negatively influenced the quality of a given study. Regarding the number of citations, manuscripts published in 2000 may have more citations because of their longer availability than those published in 2015. Therefore, publications after the introduction of ARRIVE guidelines had less time to achieve a high citation index than those published earlier. Lastly, our review included only published studies. Similar to publication bias for ischemic stroke, one may assume the same problem for preclinical SAH studies and thus negative findings are underreported.[8, 21, 22] Methodological flaws may lead to an overestimation of differences between measurements: publication bias either favors more positive results, including methodological weak studies, or dismisses the publication of negative findings that are more likely to result from methodological stronger studies.[5, 8]

In conclusion, methodological shortcomings in preclinical SAH research were prevalent throughout the 15-year study period. However, a small proportion of the published SAH experiments were methodologically very strong (MQS 8 and 9), which proves that better methodology is achievable and a realistic goal that authors should aim for. Methodological quality was neither a prerequisite for a high citation rate nor a determinant for publication in a high-impact factor journal. To overcome these shortcomings, we consider these strategies may offer the SAH research community a means to improve preclinical research quality. First, create a public platform for mandatory preregistration of all experimental investigations (analogous to clinicaltrials.gov for clinical studies). Second, support open science policies and provide repositories for storage of publicly accessible raw data. Third, strengthen efforts from journals, authors, editors, reviewers, and funding bodies to consistently follow the already existing excellent reporting guidelines. Lastly, awareness of the current situation should extend toward the grassroots researcher, including increased efforts to train the next generation in good laboratory practice. Additionally, general courses offered to all medical students is one possible initiative, and researchers who enter the field of preclinical research should consider self-learning modules and exams on handling laboratory animals and good laboratory practice.

References

1. Zoerle T, Ilodigwe DC, Wan H, Lakovic K, Sabri M, Ai J et al. Pharmacologic reduction of angiographic vasospasm in experimental subarachnoid hemorrhage: systematic review and meta-analysis. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2012;32(9):1645-58. doi:10.1038/jcbfm.2012.57.
2. Leclerc JL, Garcia JM, Diller MA, Carpenter AM, Kamat PK, Hoh BL et al. A Comparison of Pathophysiology in Humans and Rodent Models of Subarachnoid Hemorrhage. Front Mol Neurosci. 2018;11:71. doi:10.3389/fnmol.2018.00071.
3. Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. Jama. 2006;296(14):1731-2. doi:10.1001/jama.296.14.1731.
4. MacLellan CL, Paquette R, Colbourne F. A critical appraisal of experimental intracerebral hemorrhage research. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2012;32(4):612-27. doi:10.1038/jcbfm.2012.8.
5. Dirnagl U. Bench to bedside: the quest for quality in experimental stroke research. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2006;26(12):1465-78. doi:10.1038/sj.jcbfm.9600298.
6. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. PLoS Biol. 2015;13(6):e1002165. doi:10.1371/journal.pbio.1002165.
7. 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; a journal of cerebral circulation. 2008;39(10):2824-9. doi:10.1161/STROKEAHA.108.515957.
8. Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. PLoS Biol. 2010;8(3):e1000344. doi:10.1371/journal.pbio.1000344.
9. Macleod MR, O'Collins T, Horky LL, Howells DW, Donnan GA. Systematic review and metaanalysis of the efficacy of FK506 in experimental stroke. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2005;25(6):713-21. doi:10.1038/sj.jcbfm.9600064.
10. Marbacher S, Gruter B, Schopf S, Croci D, Nevzati E, D'Alonzo D et al. Systematic Review of In Vivo Animal Models of Subarachnoid Hemorrhage: Species, Standard Parameters, and Outcomes. Translational stroke research. 2018. doi:10.1007/s12975-018-0657-4.
11. Marbacher S, Fandino J, Kitchen ND. Standard intracranial in vivo animal models of delayed cerebral vasospasm. British journal of neurosurgery. 2010;24(4):415-34. doi:10.3109/02688691003746274.
12. Richter SH, Garner JP, Wurbel H. Environmental standardization: cure or cause of poor reproducibility in animal experiments? Nat Methods. 2009;6(4):257-61. doi:10.1038/nmeth.1312.

13. Marbacher S. Can Quality Improvement Tools Overcome the Translational Roadblock-the Vital Influence of the Researcher. Translational stroke research. 2017;8(3):203-5. doi:10.1007/s12975-017-0524-8.
14. 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. doi:10.1016/j.tins.2007.06.009.
15. 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; a journal of cerebral circulation. 2009;40(3):e50-2. doi:10.1161/STROKEAHA.108.525386.
16. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLoS Biology. 2010;8(6):e1000412. doi:10.1371/journal.pbio.1000412.

17. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. PloS one. 2018;13(5):e0197882. doi:10.1371/journal.pone.0197882.
18. Boltze J, Ayata C. Challenges and Controversies in Translational Stroke Research - an Introduction. Translational stroke research. 2016;7(5):355-7. doi:10.1007/s12975-016-0492-4.
19. Martic-Kehl MI, Wernery J, Folkers G, Schubiger PA. Quality of Animal Experiments in Anti-Angiogenic Cancer Drug Development--A Systematic Review. PloS one. 2015;10(9):e0137235. doi:10.1371/journal.pone.0137235.
20. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4:1. doi:10.1186/2046-4053-4-1.
21. 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. doi:10.1371/journal.pmed.1000245.
22. Reichlin TS, Vogt L, Wurbel H. The Researchers' View of Scientific Rigor-Survey on the Conduct and Reporting of In Vivo Research. PloS one. 2016;11(12):e0165999. doi:10.1371/journal.pone.0165999.
23. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nat Rev Drug Discov. 2011;10(9):712. doi:10.1038/nrd3439-c1.
24. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012;483(7391):531-3. doi:10.1038/483531a.
25. Perrin S. Preclinical research: Make mouse studies work. Nature. 2014;507(7493):423-5. doi:10.1038/507423a.
26. 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. doi:10.1007/s00415-005-0802-3.
27. Sena E, Wheble P, Sandercock P, Macleod M. Systematic review and meta-analysis of the efficacy of tirilazad in experimental stroke. Stroke; a journal of cerebral circulation. 2007;38(2):388-94. doi:10.1161/01.STR.0000254462.75851.22.
28. Baker D, Lidster K, Sottomayor A, Amor S. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. PLoS Biol. 2014;12(1):e1001756. doi:10.1371/journal.pbio.1001756.
29. Nakajima M, Date I, Takahashi K, Ninomiya Y, Asari S, Ohmoto T. Effects of aging on cerebral vasospasm after subarachnoid hemorrhage in rabbits. Stroke; a journal of cerebral circulation. 2001;32(3):620-8. doi:10.1161/01.str.32.3.620.
30. D'Abbondanza JA, Ai J, Lass E, Wan H, Brathwaite S, Tso MK et al. Robust effects of genetic background on responses to subarachnoid hemorrhage in mice. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2016;36(11):1942-54. doi:10.1177/0271678X15612489.
31. Hockel K, Trabold R, Scholler K, Torok E, Plesnila N. Impact of anesthesia on pathophysiology and mortality following subarachnoid hemorrhage in rats. Exp Transl Stroke Med. 2012;4(1):5. doi:10.1186/2040-7378-4-5.
32. Conzen C, Becker K, Albanna W, Weiss M, Bach A, Lushina N et al. The Acute Phase of Experimental Subarachnoid Hemorrhage: Intracranial Pressure Dynamics and Their Effect on Cerebral Blood Flow and Autoregulation. Translational stroke research. 2019;10(5):566-82. doi:10.1007/s12975-018-0674-3.
33. Schwartz AY, Masago A, Sehba FA, Bederson JB. Experimental models of subarachnoid hemorrhage in the rat: a refinement of the endovascular filament model. Journal of neuroscience methods. 2000;96(2):161-7. doi:10.1016/s0165-0270(00)00156-4.

34. Llovera G, Hofmann K, Roth S, Salas-Perdomo A, Ferrer-Ferrer M, Perego C et al. Results of a preclinical randomized controlled multicenter trial (pRCT): Anti-CD49d treatment for acute brain ischemia. Sci Transl Med. 2015;7(299):299ra121. doi:10.1126/scitranslmed.aaa9853.
35. Pluta RM, Hansen-Schwartz J, Dreier J, Vajkoczy P, Macdonald RL, Nishizawa S et al. Cerebral vasospasm following subarachnoid hemorrhage: time for a new world of thought. Neurological research. 2009;31(2):151-8. doi:10.1179/174313209X393564.
36. Sehba FA, Pluta RM. Aneurysmal subarachnoid hemorrhage models: do they need a fix? Stroke research and treatment. 2013;2013:615154. doi:10.1155/2013/615154.
37. 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. Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism. 2010;30(9):1619-24. doi:10.1038/jcbfm.2010.74.

Figure legends

Figure 1: Flowchart of study selection.
Numbers of studies for each animal model. After full-text screening, 70 studies were later excluded from final analysis when data extraction showed eligibility criteria were not met. (Figure slightly modified after Marbacher et al.[7]).

Figure 2: Methodological Quality Score (MQS) (10-point scale) as Function of Publication Year.
Median MQS of published studies was 4 in 2000, 6 after publication of the ARRIVE guidelines in 2010, and declined to 5 thereafter. Overall median (thick grey line), quartiles (thin line) and publication of ARRIVE guidelines (dotted orange line).

Figure 3: Adherence to desirable methodological practices defined by the MQS scale by publication year.
Itemizing the methodological quality score shows relative strict adherence by most studies to peer-review publication (99.5%) and compliance with animal welfare regulations (86.3%) but poor reporting of sample size calculation (0.9%). Reporting of random allocation to group, mortality, blinded induction of injury, and statement of potential conflicts of interest temporarily peeked after the publication of ARRIVE guidelines (marked by the dotted orange line) but later declined.

Figure 4: The Methodological Quality Score for different species separately
There is no statistically significant difference in methodological study quality noted between species.

Figure 5: Methodological Quality Score by class of anesthetics.
No anesthetic regimen was superior in terms of methodological quality. Class “barbiturates” includes phenobarbital, pentobarbital, thiopental, nembutal, thiobutabarbital, and barbiturate. Class “halogenated ether” includes halothane, isoflurane, methoxyflurane and sevoflurane, either alone or in combination with O_2 or N_2O. Class “benzodiazepines” includes diazepam and midazolam.

Figure 6: MQS of studies versus journal impact factor of publication.
Stronger study methodology reflected by a higher MQS was not associated with publication in a high impact factor journal. Boxes contain the 25%-75% quartiles, median (thick

horizontal line), and most extreme values (thin vertical line) lying with the box-edge and 1.5* the interquartile range. For readability, two studies were omitted in this graph, impact factors of 38.2 and 17.1, and methodological quality score of 6 and 3, respectively.

Figure 7: MQS of studies versus number of citations.
Stronger study methodology reflected by a higher MQS was not associated with a higher number of citations. Boxes contain the 25%-75% quartiles, median (thick horizontal line), and most extreme values (vertical dashed line) lying within the box-edge and 1.5* the interquartile range.

Tables

Table 1: Relative frequency of MQS items by species in all 765 studies.

Adherence to all 10 components that form the MQS in absolute values and percentages for each species. Peer review publication (99.5%) and compliance with animal welfare regulation (86.3%) was fulfilled by most studies over all species. In contrast, sample size calculation was only performed in 1 mouse and 6 rat studies.

	Mouse		Rat		Rabbit		Dog		Pig		Primate		Overall	
	N	%	n	%	N	%	n	%	n	%	n	%	n	%
All	56		411		168		101		4		25		765	
Peer review publication	56	100	410	99.8	167	99.4	100	99.0	4	100.0	24	96.0	761	99.5
Control of temperature	31	55.4	254	61.8	56	33.3	58	57.4	2	50.0	8	32.0	409	53.5
Random allocation to group	24	42.9	201	48.9	87	51.8	67	66.3	4	100.0	15	60.0	398	52.0
Blinded induction of injury	19	33.9	123	29.9	134	79.8	17	16.8	4	100.0	20	80.0	317	41.4
Blinded assessment of outcome	26	46.4	62	15.1	82	48.8	26	25.7	1	25.0	12	48.0	209	27.3
Monitoring of physiological variables	15	26.8	199	48.4	84	50.0	26	25.7	3	75.0	14	56.0	341	44.6
Mortality reported	27	48.2	193	47.0	54	32.1	7	6.9	0		4	16.0	285	37.3
Sample size calculation	1	1.8	6	1.5	0		0		0		0		7	0.9
Compliance with animal welfare regulations	45	80.4	354	86.1	147	87.5	86	85.1	4	100.0	24	96.0	660	86.3
Statement of potential conflict of interest	24	42.9	107	26	43	25.6	4	4.0	2	50.0	5	20.0	185	24.2

Table 2: Reporting of relevant basic animal characteristics across species.

Absolute and relative frequencies of reporting on experimental animal characteristics and use of anesthetics. These characteristics are well known to relevantly influence the severity of SAH and individual capacity of neurological regeneration, and thus experimental outcome and reproducibility, respectively.

		Strain		Sex		Age		Weight		Anesthesia	
Species	**All**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**N**	**%**
Mouse	56	49	87.5	30	53.6	22	39.3	34	60.7	45	80.4
Rat	411	386	93.9	375	91.2	28	6.8	366	89.1	380	92.5
Rabbit	168	148	88.1	127	75.6	10	6.0	146	86.9	157	93.5
Dog	101	79	78.2	57	56.4	5	5.0	92	91.1	83	82.2
Pig	4	3	75.0	1	25.0	0		4	100.0	3	75.0
Primate	25	0		5	20.0	0		11	44.0	19	76.0
All	765	665	86.9	595	77.8	65	8.5	653	85.4	687	89.8

Systematic Review and Meta-Analysis of Methodological Quality in *In Vivo* Animal Studies of Subarachnoid Hemorrhage

Basil E. Grüter (0000-0002-6314-2482), MD[1,2] , Davide Croci (0000-0002-7436-4267), MD[2] , Salome Schöpf, MD[2] , Edin Nevzati (0000-0001-8943-0290), MD[2] , Donato d'Alonzo (0000-0002-6759-2106), MD[1] , Jacqueline Lattmann, BSc[1] ,Tabitha Roth, BSc[1] , Benjamin Bircher, BSc[1] , Carl Muroi, MD[2] , Gilles Dutilh, PhD[3] , Hans Rudolf Widmer (0000-0003-3378-8765), PhD[4] Nikolaus Plesnila (0000-0001-8832-228X), MD[5] , Javier Fandino (0000-0002-5375-3087), MD[1,2] , Serge Marbacher (0000-0001-6305-7571), MD, PhD[1,2]

[1] Department of Neurosurgery, Kantonsspital Aarau, Switzerland

[2] Cerebrovascular Research Group, Department for BioMedical Research, University of Bern, Bern, Switzerland

[3] Department of Clinical Research, Clinical Trial Unit, University of Basel Hospital, Basel, Switzerland.

[4] Department of Neurosurgery, University Hospital Bern, Bern, Switzerland

[5] Institute for Stroke and Dementia Research (ISD), Ludwig-Maximilians University Munich, Germany

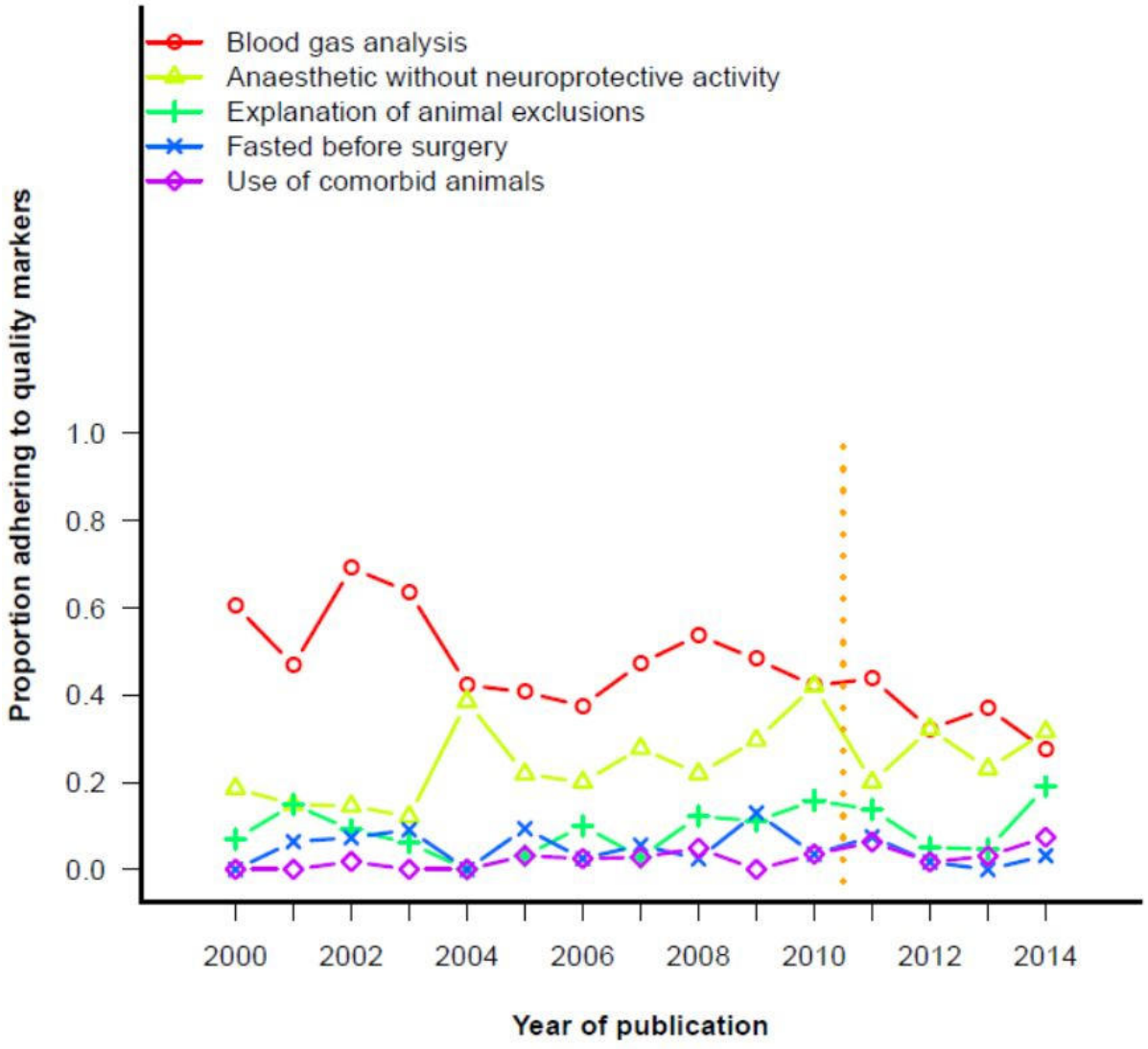
Corresponding author: Basil E. Grüter, MD, Department of Neurosurgery c/o Neuro Research Office, Kantonsspital Aarau, Tellstrasse 1, 5001 Aarau, Switzerland
+41 62 838 66 34. Email: basil.grueter@ksa.ch, neurosurgery@ksa.ch

Supplementary Materials

Supplementary Figures

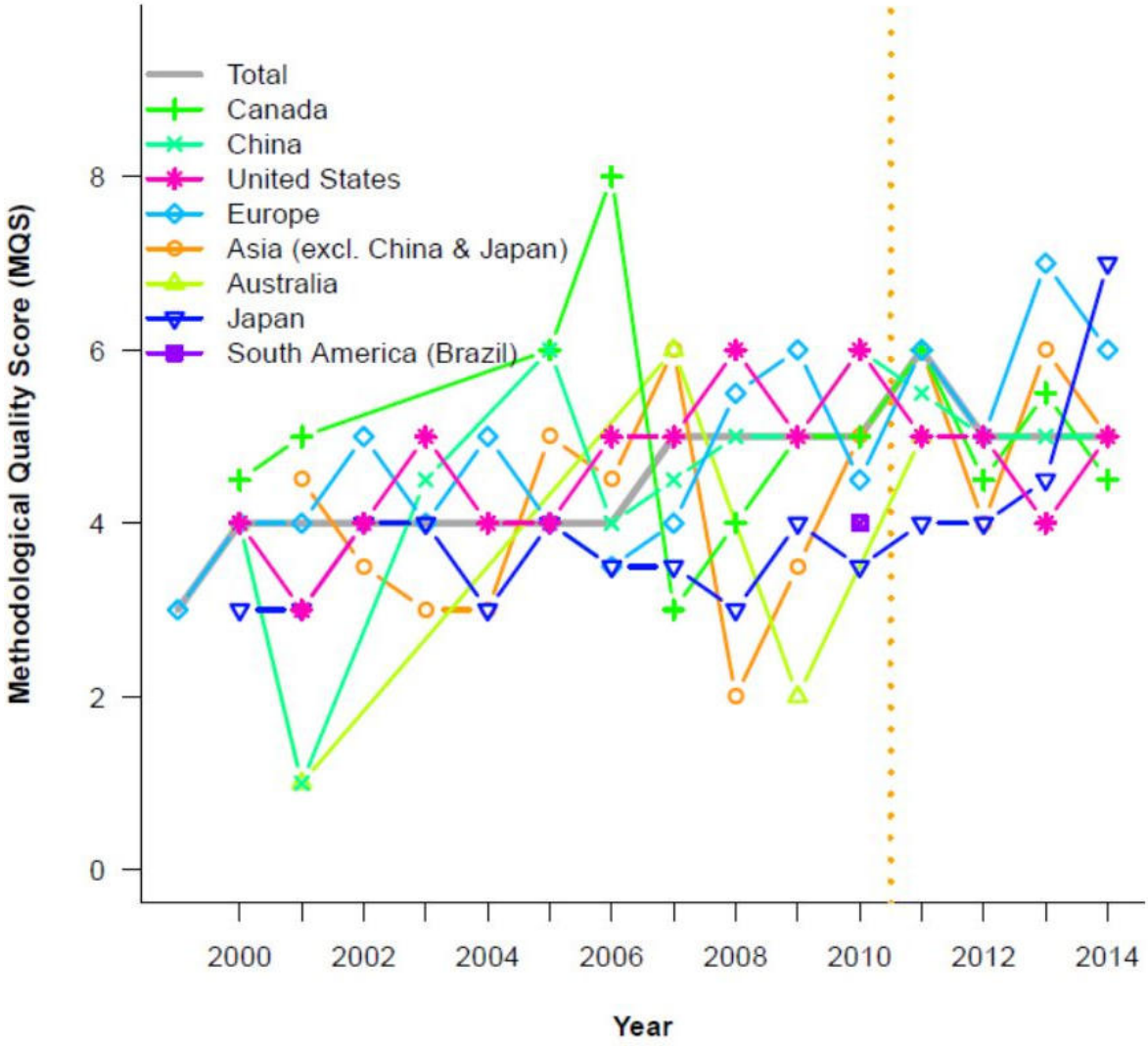
Supplementary Figure 1 – Adherence to methodological practices not included in the ARRIVE checklist or MQS.

The ARRIVE guidelines published in 2010 (dotted orange line) had no influence on frequency of reporting of methodological characteristics, not included in the ARRIVE checklist and the MQS. Note blood gas analysis, an important parameter during assessment of cerebral vascular tone, was less frequently reported over time. This trend parallels the decreased number of studies evaluating DCVS in larger animals (i.e., primates, dogs) and was less often performed in smaller animals (e.g., mice, rats).



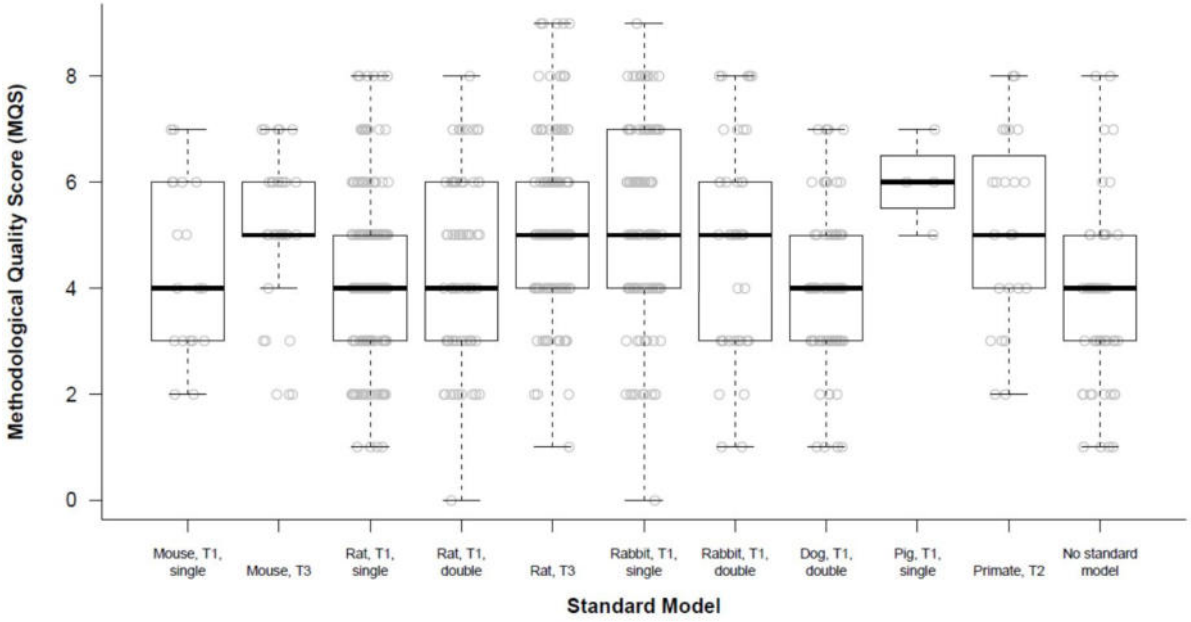
Supplementary Figure 2 – The MQS in relation to geographical origin over time.

No relevant differences in methodological quality were noted between studies by geographical origin. Dotted orange line indicates the publication year of the ARRIVE guidelines.



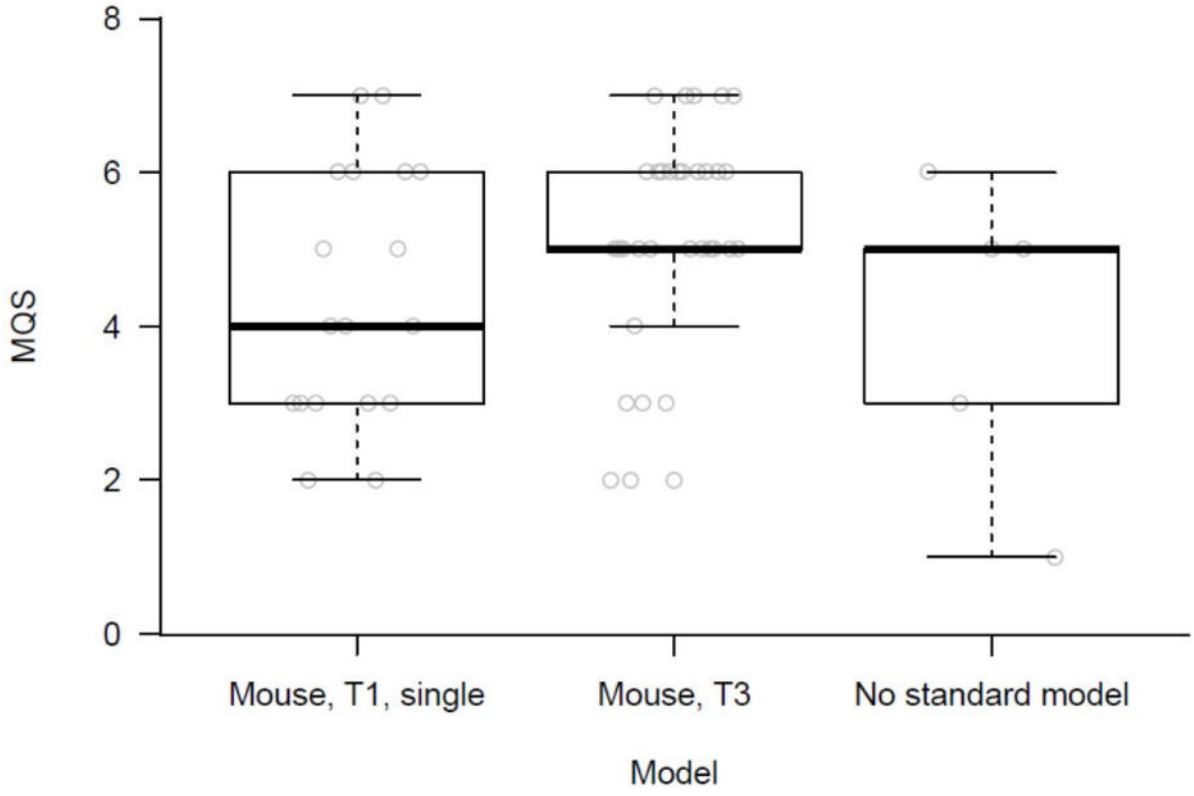
Supplementary Figure 3 – The MQS for each standard model.

No significant differences in methodological quality were noted for various standard models among species.



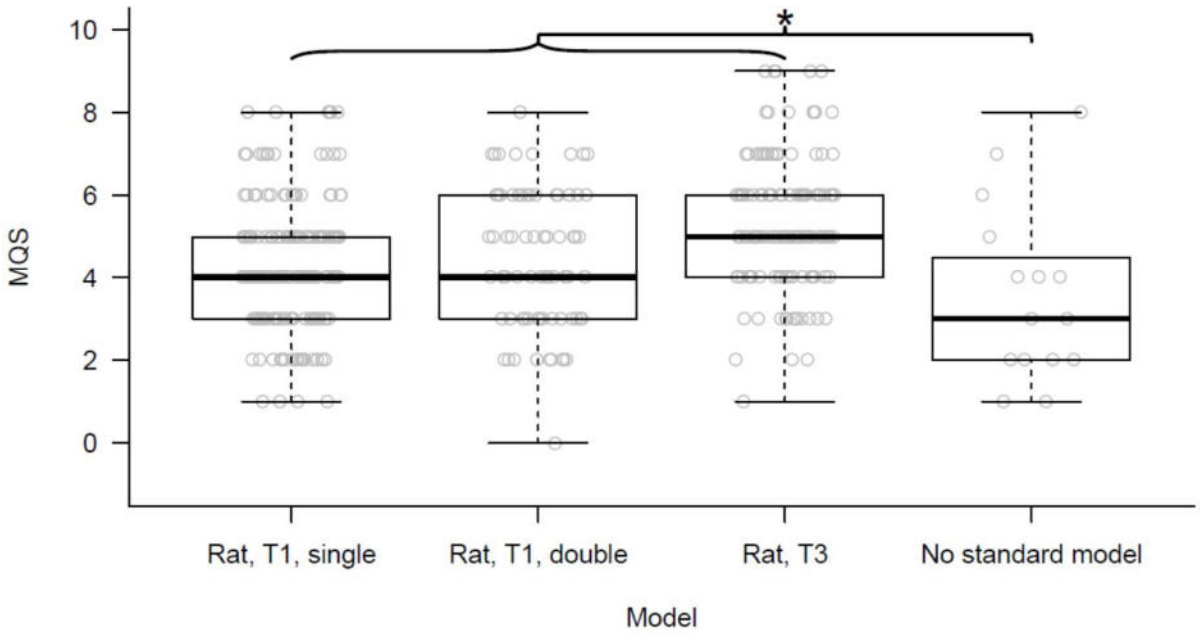
Supplementary Figure 4 – The MQS for mice studies performing standard and nonstandard models.

No significant differences between groups were noted. T1 = Single infra- and supratentorial injection of 0,06 ml autologous blood during 15 s; T3 = Endovascular perforation of a supratentorial vessel using a 5-0 suture.



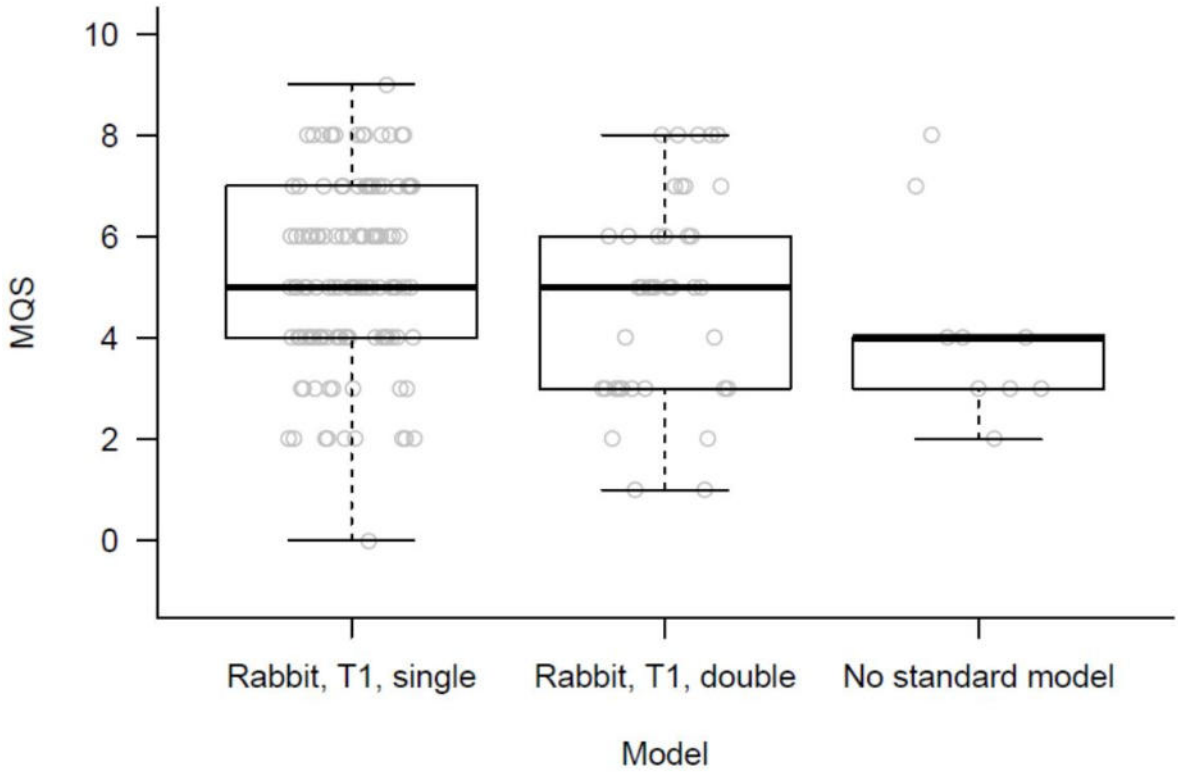
Supplementary Figure 5 – The MQS for rat studies performing standard and nonstandard models.

For rat studies, the MQS was significantly higher for studies using standard models than those using non-standard models (p = 0.018).
T1 single = Single infra- and supratentorial injection of 0,3 ml (0.1 ml/kg) autologous blood during 20 s; T1 double = Infratentorial injection of 0,3 ml (0.1 ml/kg) autologous blood during 120 s with repetition of the same procedure after 48 hours; T3 = Endovascular vessel perforation of a supratentorial vessel using a 4-0 suture.



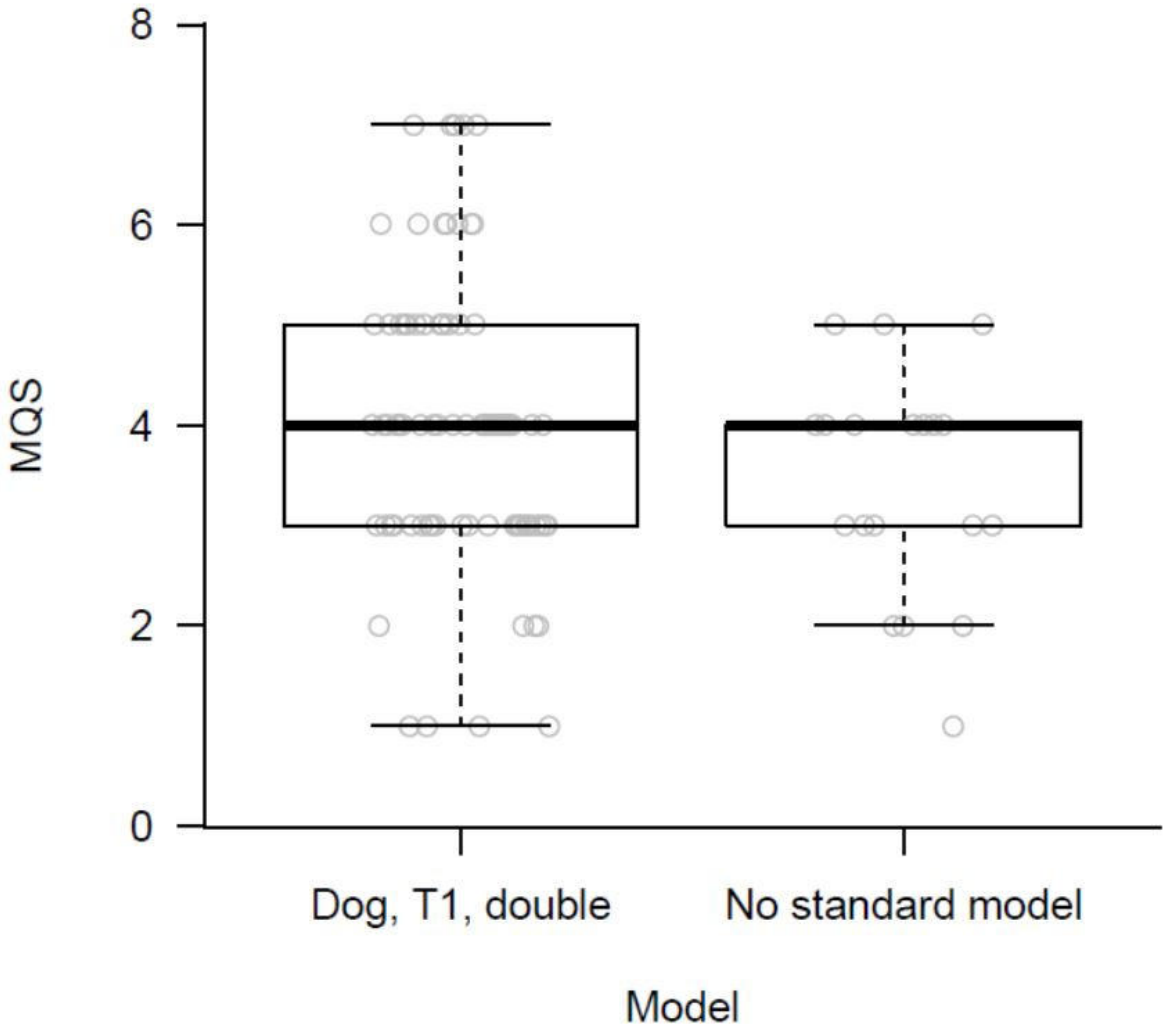
Supplementary Figure 6 – The MQS for rabbit studies studies performing standard and nonstandard models.

No significant differences between groups were noted. T1 = Infratentorial injection of 1 ml (1 ml/kg) autologous blood during 60 s; T1 double = Infratentorial injection of 1.5 ml (1.5 ml/kg) autologous blood during 60 s and repetition of the same procedure 48 hours later.



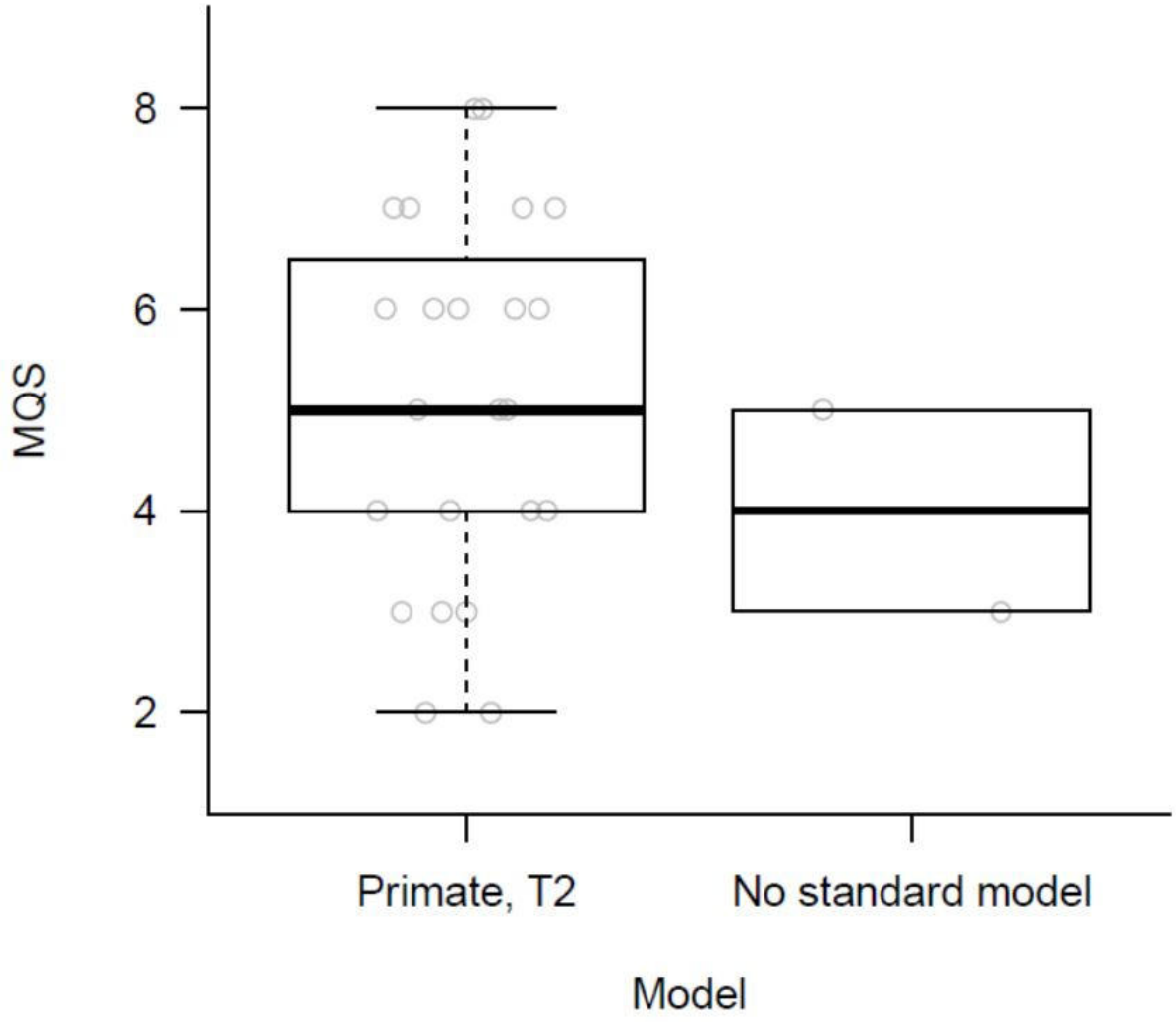
Supplementary Figure 7 – The MQS for dog studies performing standard and nonstandard models.

No significant differences between groups were noted. T1 = Infratentorial injection of 0.5 ml (0.5 ml/kg) autologous blood during 60 s and repetition of the same procedure after 48 hours.



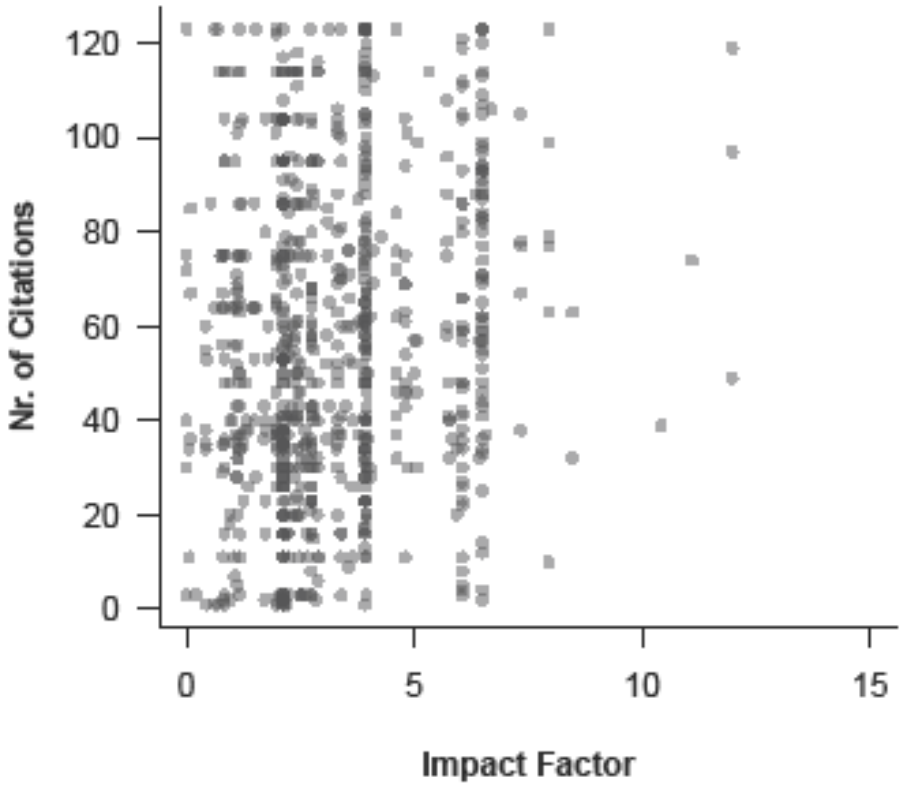
Supplementary Figure 8 – The MQS for primate studies performing standard and nonstandard models.

No significant differences between groups were noted. T2 = Craniotomy and supratentorial clot placement with 5 ml autologous blood.



Supplementary Figure 9 – Impact factor in association of number of citations per study.

The graph shows a weak ($\rho = 0.126$) but significant ($p < 0.001$) correlation between impact factor and number of citations. For readability, two studies are omitted in this graph: with impact factors of 38.2 and 17.1, and 45 and 58 citations, respectively.



Supplementary Tables

Supplementary Table 1 - Relative frequency of methodological characteristics broken down by standard model in various species.

Mouse T1: Single infra- and supratentorial injection of 0,06 ml autologous blood during 15 s; Mouse T3: Endovascular perforation of a supratentorial vessel with a 5-0 suture; Rat T1 single: Single infra- and supratentorial injection of 0,3 ml (0.1 ml/kg) autologous blood during 20 s; Rat T1 double: Infratentorial injection of 0,3 ml (0.1 ml/kg) autologous blood during 120 s with repetition of the same procedure after 48 hours; Rat T3: Endovascular vessel perforation of a supratentorial vessel with a 4-0 suture; Rabbit T1 single: Infratentorial injection of 1 ml (1 ml/kg) autologous blood during 60 s; Rabbit T1 double: Infratentorial injection of 1.5 ml (1.5ml/kg) autologous blood during 60 s and repetition of the same procedure after 48 hours; Dog T1 double: Infratentorial injection of 0.5 ml (0.5 ml/kg) autologous blood during 60 s and repetition of the same procedure after 48 hours; Pig T1 single: intracisternal blood injection, no standard model (amount of blood and time) defined; Primate T2: Craniotomy and supratentorial clot placement with 5 ml autologous blood.

	Mouse T1 single		Mouse T3		Rat T1 single		Rat T1 double		Rat T3		Rabbit T1 single		Rabbit T1 double		Dog T1 double		Pig T1 single		Primate T2		No standard model		Overall	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
all	18		33		188		71		137		120		39		82		4		23		50		765	
Peer review publication	18	100	33	100	188	100	70	98.6	137	100	119	99.2	39	100	82	100	4	100	22	95.7	49	98.0	761	99.5
Control of temperature	8	44.4	22	66.7	115	61.2	28	39.4	107	78.1	39	32.5	13	33.3	52	63.4	2	50	8	34.8	15	30.0	409	53.5
Random allocation to group	7	38.9	14	42.4	83	44.1	32	45.1	83	60.6	62	51.7	24	61.5	54	65.9	4	100	14	60.9	21	42.0	398	52.0
Blinded induction of injury	6	33.3	12	36.4	45	23.9	23	32.4	50	36.5	103	85.8	26	66.7	11	13.4	4	100	18	78.3	19	38.0	317	41.4
Blinded assessment of outcome	7	38.9	17	51.5	16	8.5	17	23.9	26	19.0	62	51.7	17	43.6	24	29.3	1	25	11	47.8	11	22.0	209	27.3
Monitoring of physiological variables	1	5.6	13	39.4	89	47.3	24	33.8	79	57.7	72	60.0	7	17.9	22	26.8	3	75	14	60.9	17	34.0	341	44.6
Mortality reported	4	22.2	21	63.6	63	33.5	35	49.3	89	65.0	37	30.8	14	35.9	7	8.5	0		4	17.4	11	22.0	285	37.3
Sample size calculation	1	5.6	0		2	1.1	1	1.4	3	2.2	0		0		0		0		0		0		7	0.9
Compliance with animal welfare regulations	16	88.9	26	78.8	165	87.8	64	90.1	116	84.7	106	88.3	34	87.2	71	89.6	4	100	22	95.7	36	72.0	660	86.3
Statement of potential conflict of interest	11	61.1	11	33.3	40	24.5	22	31.0	37	27.0	26	21.7	16	41	3	3.7	2	50	5	21.7	6	12.0	185	24.2
Anesthetic without neuroprotective activity	4	22.2	2	6.1	85	45.2	21	29.6	59	43.1	3	2.5	0		10	12.2	0		2	8.7	5	10.0	191	25.0
Use of comorbid animals	0		0		16	8.5	3	4.2	4	2.9	0		0		0		0		0		0		23	3.0
Fasted before surgery	1	5.6	3	9.1	3	1.6	1	1.4	11	8.0	11	9.2	0		0		0		2	8.7	4	8.0	36	4.7
Blood gas analysis	1	5.6	8	24.2	72	38.3	17	23.9	75	54.7	71	59.2	6	15.4	56	68.3	3	75	15	65.2	17	34.0	341	44.6
Explanation of animal exclusion	1	5.6	2	6.1	15	8.0	11	15.5	21	15.3	24	20.0	0		2	2.4	0		0		2	4.0	78	10.2

Supplementary Table 2 - Reporting of parameters of SAH induction techniques (T1 and T3) across standard models.

	Total (T1)	Injection time (T1)		Injection volume (T1)		Total (T3)	Filament size (T3)	
Species and standard model		n	%	n	%		n	%
Mouse T1 single	18	8	44.4	8	44.4	n/a	n/a	
Mouse T3	n/a			n/a		33	13	39
Rat T1 single	188	113	60.1	113	60.1	n/a	n/a	
Rat T1 double	71	31	43.7	31	43.7	n/a	n/a	
Rat T3	n/a	n/a		n/a		137	62	45
Rabbit T1 single	120	52	43.3	52	43.3	n/a	n/a	
Rabbit T1 double	39	20	51.3	20	51.3	n/a	n/a	
Dog T1 double	82	7	8.5	7	8.5	n/a	n/a	
Pig T1 single	4	3	75.0	3	75.0	n/a	n/a	
All	534	235	44.0	235	33.2	173	75	43

Mouse T1: Single infra- and supratentorial injection of 0,06 ml autologous blood during 15 s; Mouse T3: Endovascular perforation of a supratentorial vessel with a 5-0 suture; Rat T1 single: Single infra- and supratentorial injection of 0,3 ml (0.1 ml/kg) autologous blood during 20 s; Rat T1 double: Infratentorial injection of 0,3 ml (0.1 ml/kg) autologous blood during 120 s with repetition of the same procedure after 48 hours; Rat T3: Endovascular vessel perforation of a supratentorial vessel with a 4-0 suture; Rabbit T1 single: Infratentorial injection of 1 ml (1 ml/kg) autologous blood during 60 s; Rabbit T1 double: Infratentorial injection of 1.5 ml (1.5 ml/kg) during 60 s and repetition of the same procedure after 48 hours; Dog T1 double: Infratentorial injection of 0.5 ml (0.5 ml/kg) autologous blood during 60 s and repetition of the same procedure after 48 hours; Pig T1 single: intracisternal blood injection, no standard model (amount of blood and time) defined.