Aus der Klink für Neurochirurgie
der Heinrich-Heine-Universität Düsseldorf
Direktorin: Univ.-Prof. Dr. med. Katharina Faust

Meta-reviews on the reproducibility of neurosurgical preclinical research exemplified by the U-87 MG glioblastoma cell culture model and in vivo models of subarachnoid hemorrhage

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Timo Sander
2025

Als Inauguraldissertation gedruckt mit der Genehmigung der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Prof. Dr. med. Nikolaj Klöcker

Erstgutachter: Prof. Dr. med. Sajjad Muhammad

Zweitgutachter: Prof. Dr. sc. hum. Oliver Kuß

Teile dieser Arbeit wurden veröffentlicht:

Alpdogan S, **Sander T**, Zhang R, et al., 2022, Meta review on perforation model of subarachnoid hemorrhage in mice – filament material as a moderator of mortality. In German Medical Science GMS Publishing House, Conference Abstract, 73. Jahrestagung der Deutschen Gesellschaft für Neurochirurgie. (1)

Alpdogan, S*, **Sander, T***, Zhang, R et al., 2022, Meta-review on Perforation Model of Subarachnoid Hemorrhage in Mice: Filament Material as a Possible Moderator of Mortality. Transl. Stroke Res. (Vol. 15, pages 16-29), *: These authors contributed equally. (2)

Alpdogan S, Li K, **Sander T**, Cornelius JF, Muhammad S., 2023, Cisterna Magna Injection Mouse Model of Subarachnoid Hemorrhage (SAH): A Systematic Literature Review of Preclinical SAH Research. J Exp Neurol (Vol. 4(1), pages 11-20). (3)

Sander T, Ghanawi J, Wilson E, Muhammad S, MacLeod M, Kahlert U, 2022, Reproducibility of in vitro glioma research exemplified by the U-87 MG model. German Medical Science GMS Publishing House, Conference Abstract, 73. Jahrestagung der Deutschen Gesellschaft für Neurochirurgie. (4)

Sander T, Ghanawi J, Wilson E, Muhammad S, Macleod M, Kahlert DU, 2022, Meta-analysis on reporting practices as a source of heterogeneity in in vitro cancer research. BMJ Open Sci. (Vol. 6(1), pages 1-13). (5)

Zusammenfassung

Die Reproduzierbarkeit wissenschaftlicher Erkenntnisse ist eine grundlegende Bedingung für wissenschaftlichen Fortschritt. In jüngster Zeit ist die Reproduzierbarkeit jedoch insbesondere in der präklinischen Forschung in Frage gestellt worden. Da es aber nur wenige konkrete Daten über das Ausmaß von Irreproduzierbarkeit gibt, war das Ziel dieser kumulativen Arbeit, diese an insgesamt drei präklinischen Modellen aus der neurochirurgischen Forschung zu quantifizieren.

Bei diesen Modellen handelte es sich zum einen um ein konventionelles Glioblastom-Zellkulturmodell mit der weit verbreiteten, kommerziell erhältlichen Zelllinie *Uppsala-87 malignant glioma* (U-87 MG), und zum anderen um Mausmodelle der Subarachnoidalblutung, einmal mittels endovaskulärer Perforation einer intrakraniellen Arterie und einmal mittels direkter Blutinjektion in die Cisterna magna. Beiden Erkrankungen ist eine vergleichsweise schlechte Prognose beim Menschen gemeinsam, welche sich in der Vergangenheit nur geringfügig verbesserte.

Als Maß für die Irreproduzierbarkeit wurde in den jeweils modellspezifischen Meta-Reviews die Ergebnisvarianz von grundlegenden Endpunkten zwischen den Artikeln, die über das durch zufällige Streuung zu erwartende Maß hinausging, herangezogen. Nach einer systematischen Literaturrecherche, in der publizierte Originalarbeiten mit den genannten Modellen und vorab definierten Experimenten identifiziert wurden, wurden spezifische experimentelle Parameter sowie die numerischen Ergebnisse der Endpunkte in den eingeschlossenen Studien erhoben. Zur Beurteilung der Reproduzierbarkeit der Modelle wurden dann zum einen die Berichtshäufigkeiten der Parameter bestimmt, und zum anderen die Varianz der publizierten Ergebnisse derselben Experimente zwischen den wissenschaftlichen Artikeln metaanalytisch ermittelt, wobei auch versucht wurde Moderatoren der Reproduzierbarkeit zu ermitteln.

Die Literaturrecherche identifizierte zunächst 137 relevante Artikel für das Glioblastom-Modell, 42 für das Perforations- und sieben für das Injektions-Subarachnoidalblutungs-Modell, weshalb für dieses aufgrund der geringen Literaturmenge anstelle einer Meta-Analyse eine deskriptive Auswertung durchgeführt werden konnte. In vielen der Publikationen wurden teils mehrere essenzielle experimentelle Parameter nicht oder nur unklar berichtet. Dies betraf z. B. die Zellkonzentration, die Spezifikationen des für die Perforation verwendeten Filaments, sowie die Eigenschaften der Versuchstiere. Darüber hinaus wurde deutlich, dass die Modelle hinsichtlich der untersuchten Endpunkte, nämlich der Sensitivität der Glioblastomzellen gegenüber dem Standard-Chemotherapeutikum Temozolomid respektive der Mortalität nach Induktion einer Subarachnoidalblutung, anhand der publizierten Ergebnisdaten als unzureichend reproduzierbar eingestuft werden mussten. Als signifikante Moderatoren dieser Endpunkte und damit der Reproduzierbarkeit zwischen den Studien wurden die Glukosekonzentration im U-87 MG Zellkulturmedium sowie das Material des zur Perforation einer intrakraniellen Arterie verwendeten Filaments beobachtet.

Die hier vorgestellten Daten können als Grundlage für zukünftige Diskussionen zur Relevanz und Förderung der Reproduzierbarkeit in der präklinischen Forschung dienen und zeigen mögliche Verbesserungsstrategien auf. Hierbei verdeutlicht die Erschwerung der Reproduzierbarkeitsanalysen durch unzureichende Berichtshäufigkeiten vieler methodischer und ergebnisbezogener Parameter in der zugrundeliegenden Originalliteratur jedoch, dass für zukünftige Untersuchungen zunächst eine Verbesserung des wissenschaftlichen Berichtens, beispielsweise über die Entwicklung und verstärkte Anwendung von Publikationsrichtlinien, forciert werden sollte. Weitere Ansatzpunkte zur Steigerung der Reproduzierbarkeit könnten die Sensibilisierung der Autoren für die Wichtigkeit einer reproduzierbaren und transparenten Forschung sowie die Einrichtung von präklinischen Modellregistern und digitalen Austauschplattformen zur Lösung von spezifischen Reproduzierbarkeitsproblemen umfassen. Insgesamt hat die Steigerung der präklinischen Reproduzierbarkeit das Ziel, die Translationsrate von Erkenntnissen in die Klinik zu erhöhen und die wissenschaftliche Ressourcennutzung zu verbessern, was letztlich zu einem Benefit für die Patienten führen soll.

Summary

The reproducibility of scientific findings is fundamental to scientific progress. Recently, however, reproducibility has been questioned, particularly in preclinical research. As there is little concrete evidence on the extent of potential shortcomings with reproducibility, this cumulative work aimed to quantify the reproducibility of three preclinical models in neurosurgical research.

These models were a conventional glioblastoma cell culture model based on the widely used, commercially available *Uppsala-87 malignant glioma* (U-87 MG) cell line, and mouse models of subarachnoid hemorrhage (SAH) induced by endovascular perforation of an intracranial artery and by direct blood injection into the cisterna magna. Both diseases have a relatively severe prognosis in humans which has improved only slightly in the past.

In the model-specific meta-reviews, the measure of irreproducibility was the variance in results of basic model outcomes across the included articles that exceeded the expected variance due to random sampling error. Following separate systematic literature reviews to identify original research articles containing the same experiments on each of the respective preclinical models, data on their experimental parameters and numerical results were collected. Moreover, after assessing the frequency of reporting of these experimental parameters, the variance of the predefined outcomes across the articles was meta-analytically estimated whereby an attempt was made to identify drivers of reproducibility.

The literature reviews identified 137 relevant articles for the glioblastoma model, 42 for the perforation, and seven for the injection SAH model, for which the reproducibility was described descriptively because of the low number of relevant studies. For all models, a large share of articles did not include a report of several important model parameters or reported them unclearly. These parameters included cell concentrations, specifications of the filaments used for perforation, and characteristics of the experimental animals, among others. Furthermore, the analyses revealed that the models were not sufficiently reproducible concerning the investigated results, as the variance across the articles of the sensitivity of the glioblastoma cells to the standard chemotherapeutic agent temozolomide and the mortality after induction of SAH exceeded what could be statistically expected. Thereby, the glucose concentration in the cell culture medium and the material of the filament used for intracranial vessel perforation were identified as significant moderators of the outcomes across studies and therefore moderated the reproducibility.

The findings presented in this thesis can serve as a basis for future discussions on the relevance and enhancement of reproducibility in preclinical research and include potential strategies for achieving this. Thereby, the identified shortcomings in the reporting of methods and results in the primary literature compromised the meta-reviews and thus highlighted the necessary improvements for future reproducibility analyses, for example, through the development and increased implementation of specific reporting guidelines in the preclinic. Other approaches for the improvement of reproducibility may include sensitising researchers to the importance of reproducible and transparent research, as well as the establishment of preclinical model databases and digital interaction platforms for the discussion of individual issues during reproduction attempts. Overall, improving reproducibility in the preclinical setting aims to increase the translation rate of promising preclinical findings into clinical research and care, ultimately leading to improved patient outcomes.

Abbreviations

ARRIVE	Animal Research: Reporting of In Vivo Experiments
ATCC	American Type Culture Collection
BA	basilar artery
CBF	cerebral blood flow
CI	confidence interval
CNS	central nervous system
CSF	cerebrospinal fluid
CT	computed tomography
DCI	delayed cerebral ischaemia
DMEM	Dulbecco's Modified Eagle Medium
EBI	early brain injury
ECA	external carotid artery
EQUATOR	Enhancing the QUAlity and Transparency Of Health Research
FBS	fetal bovine serum
IC50	half-maximal inhibitory concentration
ICA	internal carotid artery
ICP	intracranial pressure
IDH	isocitrate dehydrogenase
JIF	journal impact factor
kg	kilogram
MCA	middle cerebral artery
mg	milligram
MGMT	O(6)-methylguanine-DNA methyltransferase
min	minutes
mmHg	millimetres of mercury (pressure)
MRI	magnetic resonance imaging
MTT	3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide
NIH	National Institutes of Health
NSC	neural stem cells
RP:CB	Reproducibility Project: Cancer Biology
RPMI 1640	Roswell Park Memorial Institute Medium 1640
SAH	subarachnoid hemorrhage
SD	standard deviation
SEM	standard error of the mean
TMZ	Temozolomide
TTF	tumour treatment fields
U-251 MG	Uppsala-251 malignant glioma
U-87 MG	Uppsala-87 malignant glioma
US	United States (of America)
WHO	World Health Organization

Outline

1 Introduction

1.1 The relevance of reproducibility in science

Reproducibility of scientific findings has increasingly become a focus of attention in the academic world, with insufficient reproducibility often cited as one of the core problems in research (6–9). Since the 2010s, even the terms “reproducibility crisis” (6, p. 1) and “replicability crisis” (6, p. 1) have been used, although its legitimacy is disputed (10). Moreover, in 2014, the National Institutes of Health (NIH) as one of the leading organisations in health research, put the strengthening of reproducibility in medical research on its agenda, highlighting the importance of reproducible research (11). In its policy, the NIH stated that progress in science is based on two main foundations (12). First, research must be carefully planned and conducted, and second, it must be conducted and reported in a reproducible manner (12). Furthermore indicating the relevance of reproducibility, an article entitled “How science goes wrong” (7, p. 1) was published on the cover of the UK newspaper *The Economist* in 2013, addressing the scientific issues resulting from an inadequate reproducibility, publication bias, and inefficient working practices to a broad audience (7). In fact, several sources from multiple perspectives and subjects have presented irreproducibility as a general problem in science (5,8,9,13–19). However, reproducibility has rarely been investigated in the field of preclinical research, where its importance is likely to be particularly high, as it, as the name indicates, precedes clinical research in human cohorts and therefore paves the way for potential future innovations in health care (8).

The reproducibility of scientific evidence is essential, as it functions as a filter and allows a self-correcting property in translational medical research (5,20,21). Reproducible results that can be confirmed in independent reproduction experiments are more robust and likely to advance from preclinical stages to clinical trials. Conversely, findings that cannot be reproduced in the preclinical stages must be examined critically and may not advance to clinical trials but must first be adequately understood and reproduced preclinically. For these filtering and self-correcting properties, the conditions for reproducibility must be met so that promising ideas are not lost to an environment not supporting reproducible research via avoidable bias and deviations in experimental protocols. Moreover, this filtering function can also be seen in the preclinical stage itself, namely in the transition from in vitro to in vivo studies. Often, drug candidates are first extensively tested in cell model screenings, and afterwards, if positive and reproducible results were observed, more precise data on efficacy, dosage, and toxicity are obtained in animal models. Reproducibility can also be seen as a prerequisite for the optimal utilisation of available resources, especially regarding experimental animals, and thus for acting in accordance with the so called “principles of the 3Rs [sic]” (22, p. 1) in animal studies, which stand for “Replacement, Reduction and Refinement [sic]” (22, p. 1).

Furthermore, reproducibility in the form of a comprehensive and transparent reporting of experimental methods and results facilitates the interaction between scientific groups in which they frequently benefit from external published research, for example by both positive and negative findings as an input for their own research, ideally leading to an efficient collaboration of experts around the

globe (5,23). Conversely, incomplete and untransparent reporting can reduce the quality of published science (24). For example, the interpretation of reported results may be avoidably complicated because of the inability to estimate the impact of confounders arising from the individual experimental design. This could, for instance, result in an overestimation of the investigated effect and, in the groundwork area of preclinical research, to clinical trials based on incorrectly assumed conclusions (14,25). Furthermore, incomplete reporting may prevent other researchers from performing subsequent studies building on the published findings because of an increased difficulty due to an unclear previous with reduced chances to confirm the preceding findings due to a high risk of unintended study deviations.

Given these fundamental advantages of reproducible and well-reported research, it is reasonable to assume that these quality attributes must be met in preclinical research, since this constitutes the basis of biomedical research. However, as there is currently limited evidence on the extent of irreproducibility and reporting deficits particularly in preclinical research, this thesis objectifies these quality criteria in selected preclinical neurosurgical experimental models.

1.1.1 Definition of reproducibility

To answer the question of how reproducible scientific findings are, it is important to clearly define the term reproducibility, as reproducibility is understood by some as a "catch-all" (13, p. 4) phrase covering multiple shortcomings of research without a clear definition of the concrete problems. To distinguish between the terms repeatability, replicability, and reproducibility, which are often used synonymously, the following proposal has been made (26). Repeatability means that the same group of researchers obtains the same results using the same experimental setup (26). However, replicability means that different groups of researchers obtain the same results using the same experimental setup (26). Reproducibility, in contrast, means that different research groups with different experimental setups achieve the same results or conclusions (26).

In another interpretation, the term reproducibility is further subdivided into "methods reproducibility, results reproducibility, and inferential reproducibility" (27, p. 1). Therein, "methods reproducibility" (27, p. 1) addresses whether the experimental setup is described in sufficient detail to allow an reproduction of the experiment by others, regardless of the results. In contrast, "results reproducibility" (27, p. 1) refers to whether the numerical results can be reproduced while "inferential reproducibility" (27, p. 1) indicates whether the same conclusions can be drawn in a repeated performance of a published experiment. Similar to "methods reproducibility" (27, p. 1) is the term "preproducibility" (28, p. 1), which also describes whether the methodology is described in sufficient detail to allow a methodological reproduction of an experiment regardless of the result.

In this thesis, the “preproducibility” (28, p. 1) and "methods reproducibility" (27, p. 1) were investigated through the completeness of reporting of key experimental parameters, but are not referred to as reproducibility, but instead are named methodological reporting quality to minimise the risk of confusion with the “results reproducibility” (27, p. 1). Moreover, in this thesis, the term reproducibility is used to refer to the "results reproducibility" (27, p. 1) unless explicitly stated otherwise, as the

numerical results from published experiments were extracted and processed meta-analytically for the quantification of reproducibility.

1.1.2 Past attempts at measuring the extent of irreproducibility in the preclinic

Although the availability of concrete data on the prevalence and extent of irreproducibility is not extensive, attempts have been made to determine the dimensions of this problem.

In 2011, researchers from the haemato-oncology division of the biotechnology company *Amgen* (Thousand Oaks, California, United States (of America) (US)) attempted to reproduce the findings of 53 so-called "landmark" (8, p. 2) studies. Surprisingly, this was successful in only six cases, resulting in a reproducibility failure rate of 88.7 %. Despite criticism regarding the non-transparent selection of the investigated studies (29), their work suggested the relevance of limited reproducibility.

Furthermore, the *Reproducibility Project: Cancer Biology* (RP:CB), initiated by the journal *eLife*, collected reproduction studies of cancer biology experiments published in frequently cited journals (20). Eventually, of the initially targeted 193 experiments from 53 articles, reproduction studies could only be performed for 50 experiments from 23 articles due to unclear method descriptions (20). In these 50 experiments, the rate of successfully reproduced experimental results was 45.5 % (20). With these insights, the RP:CB highlighted the reproducibility barrier of insufficiently detailed descriptions of experimental methods (20). However, the RP:CB was criticised for a potential underestimation of reproducibility due to the exclusion of more complex methods (such as difficult mouse models or antibody-based immunostaining techniques) and the use of deviating reagents and measurement methods (e.g. flow cytometry instead of quantitative real-time polymerase chain reaction) (30).

Moreover, in a 2016 survey among 1576 scientists from a wide range of research disciplines, almost 90 % of respondents agreed that the term reproducibility crisis applies to the current situation in science (31). A major reason given for this subjectively perceived lack of reproducibility was the absence of accurate method descriptions (31). Interestingly, only 41 % of researchers reported that they have already implemented concrete measures to improve reproducibility in their own laboratories (31). In a similar survey among oncologic researchers, 58 % of 263 respondents reported that they have at least once experienced failure to reproduce published results (32). Confirming such a high rate of reproducibility issues, in a survey by the *American Society for Cell Biology*, 72 % of the 869 participants reported that they have been unable to reproduce published results at least once (13). Additionally, about a third of them has been informed at least once that their own results could not be reproduced by other researchers, while the most common approach for resolving reproducibility shortcomings was direct communication between researchers (13).

1.1.3 Postulated reasons for insufficient reproducibility

The causes of irreproducibility are likely to be multifaceted and often interrelated. However, the following factors in particular have been proposed to limit reproducible research results. Overall, the arguably most frequently mentioned reason for insufficient reproducibility were deficiencies in the description of the applied methods, which lead to uncertainties in the attempt of reproducing published results (24).

Furthermore, non-validated experimental reagents may contribute to irreproducibility, as their purity and correct composure are essential elements of any experiment (13,16,33). This applies to chemicals, antibodies, and nucleic acids, as well as the genetic identity of cell lines and species of laboratory animals, among others (33). In addition, it is important to regularly calibrate experimental equipment, including measuring instruments, to avoid systematic errors in data collection (33).

Moreover, the intrinsic heterogeneity of biological models themselves may also contribute to heterogeneous results as the complexity of biological processes is often greatly simplified to keep models practical. Such simplifications may result in uncontrolled important parameters of the modelled diseases that uncontrollably moderate outcomes, for example at subcellular processes. It is also disputed whether the generalisation of results obtained using a single experimental model may be a general misconception, as one particular model under standardised experimental conditions probably not allows such a generalisation because of an inadequate reflection of the heterogeneity of studied diseases (34).

Another suspected reason for insufficient reproducibility is the scientific culture, which is often pressured by financial and career incentives to submit new, positive, and innovative, sometimes even sensational results, as these are more likely to be included in highly cited journals (34,35). This could lead to the known phenomenon of publication bias where positive and desired results are much more likely to get published than negative results which would unilaterally restrict the dispersion of research results (27). In addition, selective reporting of conducted experiments may prevent other research groups from reproducing published results as they may be valid solely under very specific experimental settings (27).

Other factors potentially driving irreproducibility may include varying degrees of training and experience of researchers (33), inappropriate study designs (including underpowered sample sizes and suboptimal selected control groups) (33), and the unavailability of transparent results (36). In addition to these known or at least postulated causes of irreproducibility in science, it is likely that additional unknown and model-specific causes of irreproducibility exist.

1.1.4 Current initiatives to improve reproducibility

As the awareness of the potential consequences of non-reproducible scientific findings has progressively developed, several initiatives have been implemented to strengthen reproducibility.

The main strategy is the stronger implementation of reporting guidelines to increase “methods reproducibility” (27, p. 1) (24). However, the availability and use of such reporting guidelines is still sparse in the preclinical stage (37), as demonstrated by the practically nonexistent listing of reporting guidelines in the *Enhancing the QUAlity and Transparency Of Health Research* (EQUATOR) network, an overview platform of reporting guidelines for different scientific fields (38).

In contrast, reporting guidelines are more common in vivo studies (38), where the probably most widely used guideline is the *Animal Research: Reporting of In Vivo Experiments* (ARRIVE) guideline, which was established in 2010 to improve the quality of methodological reporting and thus the reproducibility, credibility, and usability of information gained from in vivo studies (39). In the updated version released ten years later, particular emphasis was placed on open data availability and

prioritisation within the 21 reporting items (40). According to it, the following parameters should be reported in every in vivo study publication: study design, sample size calculation for the needed number of animals, inclusion and exclusion criteria for experimental animals, randomisation, blinding, outcome assessment methods, statistical analysis procedures, characteristics of experimental animals, a detailed description of the experimental setting and procedures as well as the complete results of all conducted experiments (40). Additionally, the following parameters are recommended but not considered essential to declare: availability of a concise abstract, presentation of relevant background information and study objectives, ethics statement, details of animal husbandry, care and monitoring, interpretation of results in terms of scientific conclusions and their generalisability, statement regarding study protocol registration, and a conflict of interest statement (40). During the selection of the parameters investigated in this thesis, the items included in the ARRIVE guideline were considered to analyse the reporting quality and possible influence on the reproducibility of the animal models (2,3,40). Moreover, there are more specific reporting guidelines for in vivo neurovascular stroke research, such as the *Stroke Therapy Academic Industry Roundtable* standards (41) and the *Stem Cell Therapies as an Emerging Paradigm in Stroke* recommendations (42). Both specifically emphasise the importance of randomisation, blinding, sample size calculation, and the confirmation of findings, e.g. in another experimental animal species or in independent external laboratories (41,42). In addition, the *RIGOR guidelines* have been developed to improve the translation of preclinical research into clinical application by requiring a precise description of the experimental design, including a justification for the appropriateness of the control group and chosen model, steps taken to reduce potential bias, and a critical discussion of the clinical relevance and reproducibility of the study (43). However, despite the existence of such reporting guidelines for in vivo studies, reviews have suggested that the implementation appears to be not wide enough to significantly improve reporting practices (44,45).

As the availability and use of reporting guidelines in the in vitro field is even lower and less standardised, reporting in this field has even been described as “Wild West” (37, p. 1) and “anything goes” (37, p. 1). Indeed, the EQUATOR network does not contain any broad reporting guidelines specifically for in vitro research as of June 2024 (38). If authors nevertheless prefer to adapt their reporting based on predefined criteria, they may adhere to more broad guidelines for experimental randomised controlled trials, although these were primarily designed for clinical trials (38). However, for specific areas of preclinic research with complex experimental models, more detailed model-specific instructions are needed (46). Therefore, guidelines for laboratory practice itself that also include recommendations for the publication of results may be consulted (37). For example, the *Guidance Document on Good In Vitro Method Practices* has been published with the overall aim of ensuring a minimum standard quality of in vitro research, particularly in cell culture experiments, including their presentation (47). Overall, however, there is also a clear need for improvement in in vitro reporting, as numerous investigations have demonstrated (48–52).

Besides the implementation of reporting guidelines as the main strategy for improving the preclinical research reproducibility, several open access and best practice initiatives have been launched

to facilitate reproducibility and transparency (53,54). The public registration of study protocols developed prior to the conduction of studies, including details on planned experiments, in- and exclusion criteria, study endpoints, analytical methods, and research hypotheses, has been promoted to reduce the risk of bias during the publication processes due to selective reporting (55).

However, overall, additional and more diverse approaches are probably necessary for a sustained enhancement of reproducibility.

1.2 Investigation of reproducibility in selected neurosurgical preclinical models

To quantify the reproducibility of preclinical research, this thesis investigated three models of two different diseases of the central nervous system (CNS). The first disease is the highly malignant neoplasia glioblastoma, modelled by the in vitro cell culture of the commercially available *Uppsala-87 malignant glioma* (U-87 MG) cell line (5). The second disease is the neurovascular emergency of subarachnoid hemorrhage (SAH), modelled in vivo by the endovascular perforation mouse model (2). In addition, a concise comparison with the cisterna magna blood injection SAH mouse model was performed (3).

One might ask how these two very different models fit into a joint research project since they compare two diseases with highly different aetiology, pathogenesis, and treatment, moreover, modelled in both in vitro and in vivo conditions. In fact, it was actually this very disparity between the models and diseases that motivated their inclusion in a comparative analysis. Since these two complementary models were likely to have different model-specific requirements for reproducibility, the aim was to analyse whether there are nevertheless similarities in the reporting qualities as well as the reproducibility and its moderators. If such similarities would be identified, the generalisability of conclusions about reproducibility in preclinical research would be higher than if only one model or similar models of the same disease would have been investigated. In addition, the inclusion of both an in vitro and an in vivo model covered the two most common preclinical model types, which may further increase generalisability. Moreover, all examined models are well established in their respective fields and the PhD student's supervising research group (56,57). Furthermore, the investigated models had in common that they are often used to test potential therapeutic approaches for the two diseases, which both have a poor prognosis with only moderate improvements in recent years (58,59). However, in order to maintain the required independence, the individual models were investigated in separate meta-analyses, each with distinct model-specific parameters and outcomes of interest.

1.2.1 U-87 MG in vitro glioblastoma model

The U-87 MG glioblastoma model is based on an in vitro cell culture of the commercially available U-87 MG cell line (60). It was derived from a 44-year-old human patient with glioblastoma at Uppsala University in 1966 (61). It has been found, however, that the current samples of U-87 MG, offered by the American Type Culture Collection (ATCC) (Manassas, Virginia, USA), differ significantly from the original version in their genetic identity, leaving the derivation unclear (62,63). Nevertheless, the U-87 MG model is still widely used and, like other cell line models for in vitro cancer

research, has the advantages of relatively low cost, a good practicability, and the possibility of genetic alterations for molecular pathway studies (30).

1.2.1.1 Glioblastoma

Glioblastoma is a highly malignant neoplasia and marks the second most common CNS-derived tumour behind meningioma (64,65). An analysis of CNS tumours in the US population from 2014 to 2018 showed an incidence rate of approximately 3.2 per 100,000 people per year (66). There, men were 1.6 times more frequently diagnosed with glioblastoma than women (66). The median age at diagnosis was 65 years (66). Due to the relatively high incidence of glioblastomata in younger patients and the short life expectancy after diagnosis, glioblastoma is one of the malignancies with the most active patient years lost (67).

Although presenting relatively unspecific, common symptoms of glioblastoma include headache, nausea and vomiting, seizures, changes in personality, mood, memory or concentration, and localised neurological problems, such as motor or sensory deficits, depending on the location of the tumour (68). Glioblastomata are most commonly found in the frontal (about 40 %) and temporal (about 30 %) lobe (69). In addition to a detailed physical examination with special attention to focal neurological deficits and signs of increased intracranial pressure (ICP), contrast-enhanced magnetic resonance imaging (MRI) or computed tomography (CT) scans may raise the suspicion of glioblastoma (68). However, a definitive diagnosis, including subtyping, requires a molecular pathological histological examination of biopsy material or intraoperatively removed tumour tissue (68). In the most recent 5th edition of the World Health Organization (WHO) classification of CNS tumours, the term glioblastoma was solely assigned for isocitrate dehydrogenase (IDH) wild-type diffuse gliomas (70). The IDH mutant glioblastomata according to the previous classification were reclassified as IDH mutant diffuse gliomas (70). With WHO grade IV, glioblastomata belong to the most aggressive group of CNS neoplasias with the worst prognosis (70).

The current first-line treatment of glioblastoma is a multimodal approach including surgical resection if feasible, followed by adjuvant radiotherapy and chemotherapy containing the alkylating agent temozolomide (TMZ) (71,72). In younger patients in relatively good condition, lomustine as an additional alkylans may be added in the presence of O(6)-methylguanine-DNA methyltransferase (MGMT) promoter methylation (73). Moreover, tumour treatment fields (TTF) may be considered as an additional therapeutic option in adult patients (74). As side effects caused by either the diffuse expansion of the tumour, peritumoural edema or the treatment, seizures and other neurological impairments are often managed with levetiracetam and dexamethasone, respectively (75).

The prognosis based on survival data of glioblastoma patients in the US is very poor, with a median overall survival of approximately ten months after diagnosis (58). Thereby, both surgical resection (14.9 months) and radiochemotherapy (16.9 months) significantly improve survival (58). The methylation of the MGMT promoter and younger patient age are independent positive prognostic factors (58). However, long-term survival is very limited, with a three-year survival rate of less than five percent and only occasional reports of survival beyond five years after diagnosis (76), so advances in therapy

are urgently needed. Promising prospects for such advances include, among others, immunotherapy, synthetic molecules (e.g. RES-529, ATX-101, GLPG1790), and natural compounds (e.g. trans-sodium crocetinate, oleandrin) (77).

1.2.1.2 Variables within the U-87 MG model

In experiments using the U-87 MG glioblastoma model, some parameters may significantly influence the properties of the cell culture and thus the response to tested treatments. Therefore, the potential influence of these parameters on the reproducibility of TMZ sensitivity was meta-analytically assessed in this thesis, as these model parameters can vary between studies.

First, since, as mentioned above, the genetic identity of cell lines can change due to genetic drift, shift, and the accumulation of chromosomal aberrations over time, as it has been observed in the U-87 MG cell line (62,78). Therefore, conducting a cell line validity check has been recommended to verify the appropriateness of the model, e.g. by sequencing short tandem repeat regions (79–81). Moreover, the cell passage at which experiments with U-87 MG cells are performed may also influence the characteristics of the in vitro model. With increasing numbers of cell passages, cells with a higher proliferation rate due to genetic alterations may dominate the overall tumour cell population (82). Furthermore, cell density may also modulate the characteristics of the model due to inter-cell-communication via released mediators and a variating availability of nutrients. Indeed, a relationship between cell density and treatment response has been shown for several cell lines, interestingly with both increased and decreased drug sensitivities at higher densities (83,84).

Equally important to the characteristics of a cell culture is the culture medium with its additives (47). In a 2017 review, Dulbecco's Modified Eagle Medium (DMEM) and Roswell Park Memorial Institute Medium 1640 (RPMI 1640) were listed as by far the most frequently used cell culture media in in vitro research (85). Interestingly, they differ widely in their composition, including the concentration of phosphate, calcium, magnesium, vitamins, and amino acids (85,86). As a general trend, DMEM tends to promote growth in adherent cultures, whereas RPMI 1640 tends to favour the establishment of suspension cultures (86). To avoid confounding by different culture media in the meta-analysis on the reproducibility of the U-87 MG model, this thesis focused exclusively on DMEM as the culture medium because a previous review identified it as the most commonly used culture medium for in vitro glioblastoma models (87). However, even when using the same basic culture medium, differences in its exact composition are possible due to different concentrations of certain ingredients such as glucose, antibiotics, and fetal bovine serum (FBS). Especially the availability of glucose can affect cell metabolism and alter the sensitivity of U-87 MG cells to TMZ treatment (88,89). Similarly, the addition of antibiotics has been observed to influence cancer cell differentiation (90). Likewise, serum supplementation, mainly with FBS, containing hormones and nutrients is important for the proliferation characteristics of cell cultures (91). Due to its variation, it is recommended that FBS should solely be obtained from certified distributors with a documentation of the source (47,92). In addition, mycoplasma contamination can alter the genetic and phenotypic characteristics of cell cultures, so regular screenings are recommended (17,47,93).

The investigated outcome for the reproducibility assessment was the inhibition of cell viability by the first-line chemotherapeutic agent TMZ compared to the control group not receiving TMZ as a commonly measured indicator of standard treatment response allowing an extensive analysis across a large body of published literature (5).

1.2.1.3 Comparison with other in vitro glioblastoma models

Besides U-87 MG, there are also other well-established so-called classical immortalised cell lines that are cultured under serum conditions (60), for example *Uppsala-251 malignant glioma* (U-251 MG), T98G and LN-229 (60,94). As a result of these classical cell lines not optimally reflecting the stem cell characteristics of glioblastoma cells (95), newer models emerged based on neural stem cells (NSC). In these, a key difference from classical cell lines is the serum-free growth (60). NSC models, including those obtained from individual patient tumour tissue, can be cultured in either adherent or suspension conditions (96,97). Adherent culturing has been recommended as it better promotes stem cell properties and thus more accurately models tumour heterogeneity with the ability of self-renewal and differentiation (95). Inadequate coverage of these characteristics in the classical cell lines has been considered as one factor contributing to difficulties in translating findings to the clinic (60). Nonetheless, questions have also been raised regarding the translatability of the newer NSC models, as quiescent cells, which are thought to be important for therapy resistance mechanisms, can be replaced by cell subtypes with higher proliferation rates and therefore may be underrepresented (98,99).

1.2.2 Endovascular perforation SAH mouse model

In contrast to in vitro glioblastoma models, studying SAH in living animals poses very different demands on researchers. For this purpose, scientists can choose between a variety of in vivo models to address specific questions in the complexity of the disease. In fact, a review identified 72 different animal models of SAH (100). Of those, commonly used models involved either the injection of autologous blood into the subarachnoid space or the perforation of an artery by an endovascularly introduced filament, resulting in extravasation of blood into the subarachnoid space (100). Thereby, a variety of experimental animals, from rabbits to primates, have been used, but as mice were predominantly chosen (43), this review focused on them as experimental animals to ensure sufficient comparability within the analysis. Likewise, this thesis focused on the endovascular perforation model as the most frequently used in vivo SAH model among all mouse models and included a comparison with an also commonly chosen injection model (2,3,45).

1.2.2.1 Subarachnoid hemorrhage (SAH)

A SAH is defined as a bleeding into the subarachnoid space, located between the arachnoid and the pia mater. (101). It can be caused by trauma or occur spontaneously, although traumatic genesis is more common (101). In spontaneous SAH, rupture of a preexisting aneurysm is the cause in approximately 85 % of cases (102). A further 10 % of spontaneous SAH is caused by perimesencephalic nonaneurysmal bleedings while the remaining 5 % have various causes (103). The aneurysms in the aneurysmatic spontaneous SAH are most commonly of the saccular type and develop as a result of degeneration of the internal elastic lamina under haemodynamic stress leading to thinning and loss of

the muscular tunica media (103). Risk factors for aneurysm development include arterial hypertension, smoking, alcoholism, cocaine consumption, and family history (101). SAH has a global average incidence of about 9 per 100,000 persons per year (104). However, the incidence varies regionally with Finland and Japan having reported the highest rates at about 20 per 100,000 persons per year (104). Overall, women have been reported having an approximately 1.6 times higher risk of SAH than men (105). The average age of patients with SAH was between fifty and sixty years (103,106–108).

The classic description of the leading presenting symptom of SAH is the "thunderclap headache" (109, p. 1), which is a sudden onset of a very severe head pain (110). Accompanying symptoms may include nausea and emesis, neck pain, seizures, and a reduction or loss of consciousness (110). Thereby, the onset of symptoms is often associated with physical exertion (111). However, up to half of SAH patients present without clinical signs in the neurological examination (112). The diagnostic algorithm includes a native CT of the cranium and an optional lumbar puncture if the CT is inconclusive for SAH and there are no signs of an increased ICP (112,113). Regarding CT scans, the Fisher score grades the amount of blood in the subarachnoid space from grade I with no visible bleeding to grade IV with diffuse bleeding (114). The grading serves as a predictor for vasospasm after SAH, while the highest prevalence of vasospastic complications is in grade III (115). Critical for the outcome of patients suffering SAH is a prompt initiation of diagnostics and treatment, as SAH can rapidly lead to failure of vital brain functions due to a decrease in cerebral blood flow (CBF) and resultant cerebral ischaemia, as well as direct damage to brain tissue due to the mass effect of the bleeding (116). Further early complications of SAH include rebleedings, hydrocephali, seizures, and elevated ICP with subsequent loss of consciousness (117). These complications occur quite often, as rebleedings have been observed in approximately 15 % of SAH patients (118), hydrocephali in approximately 50 % (116), and seizures in approximately 25 % (119). Additionally, the leading complication of SAH in the subacute phase is the phenomenon of early brain injury (EBI) (117). The concept of EBI was introduced in 2004 (120) and comprises the pathophysiological processes that occur within the first 72 hours after SAH (121). It is recognised as a major contributor to the high morbidity and mortality of SAH (122) and has a multifactorial pathophysiology with a decreased cerebral perfusion due to increased ICP (121), an impaired cerebral tissue autoregulation (123), edema (124), blood-brain barrier impairment (125), microthrombosis (122), oxidative stress (122), ischaemia (122), and tissue irritation due to the subarachnoid blood (126). Moreover, in the following period after SAH, the phenomenon of delayed cerebral ischaemia (DCI) is the most important complication in survivors of the subacute phase (117). DCI is defined as a neurological deterioration lasting at least one hour (either a focal neurological deficit or a reduction of at least two points on the Glasgow Coma Scale) that cannot be explained by an additionally present disease and is detectable on brain imaging with evidence of new ischaemia or infarction between 48 hours and six weeks after the onset of SAH (127). The earlier hypothesis of vasospasm as the direct cause of DCI has been replaced by a more complex theory, in which DCI and vasospasm may result from the same mechanisms, including vascular dysfunction, microthrombosis, neuroinflammation, and cortical spreading depolarisations (127). However, vasospasm is still

recognised as a clinical indicator of DCI (127). In addition, SAH can lead to a systemic inflammatory response syndrome, affecting pulmonary, cardiac, homeostatic, and other organ system functions (128).

Treatment of SAH includes basic measures such as bed rest, adjustment of mean arterial pressure to 60 to 90 mmHg, analgesia, antiemetics, and the maintenance of neuroprotective conditions such as normovolemia, normoglycemia, and normothermia (116). These interventions are primarily intended to reduce the risk of recurrent bleeding and secondary vasospasm (116). Additionally, the calcium channel antagonist nimodipine should be administered for the first 21 days after SAH for vasospasm prevention (116). Furthermore, in cases of aneurysmal SAH, endovascular coiling or surgical clipping of the affected artery is recommended and should be performed as soon as possible if feasible (116). Also, epileptic seizures may require anticonvulsant therapy, such as levetiracetam (116). A possible hydrocephalus may require depressurization by an external ventricular drainage or a cerebrospinal fluid (CSF) shunt (116). The 30-day mortality rate of patients with SAH has been reported between 25 and 35 % (129–131), while the 90-day mortality rate was between 35 and 40 % (103). Thereby, the overall pre-hospital mortality was high at around 15 % of patients dying before reaching the emergency department (132). Unfortunately, mortality rates have decreased only slightly in recent decades so that the prognosis remains poor (131). In addition, more than half of survivors experience inadequate recovery of their neuropsychiatric health status (103) and approximately one third remain severely disabled (133). Although SAH has a lower incidence rate than strokes caused by ischaemia and intracerebral hemorrhage, it is a highly relevant disease in terms of active years of life lost due to the relatively younger age of patients and high morbidity and mortality (134).

1.2.2.2 Endovascular perforation procedure

A popular method of studying SAH is the in vivo endovascular perforation technique. It was first described by Barry and colleagues, who, after removing the skull, punctured the basilar artery (BA) in rats which led to an extravasation of blood (135). As they microscopically detected the presence of blood in the subarachnoid space and successfully monitored vasospasm over the following days, they demonstrated the applicability of the model for SAH studies (135). However, in the currently most widely used variant of this model, an intracerebral artery is perforated with an intravasally introduced filament, which has the advantage of reducing surgical trauma to the animals (136,137). The SAH induction surgery requires sufficient anaesthesia of the experimental animal. When selecting mice as experimental animals, this can be realised, for example, by inhalational induction with 4 % isoflurane and 30 % oxygen, followed by maintenance with intraperitoneally injected fentanyl (0.05 mg/kg), medetomidine (0.5 mg/kg), and midazolam (5 mg/kg) (136,138). Then, the neck is exposed in a supine position and the skin is opened medianly (136,137). The underlying tissue between the salivary glands is then incised to expose the large cervical arteries with their branches (136,137). Next, the external carotid artery (ECA) is ligated with a suture (136,137). After microclipping the common and internal carotid artery (ICA), a filament is inserted into the ECA and secured with another suture (136,137). After removing the microclips, the filament is pushed through the ICA towards the brainstem with the aim of perforating an artery of the circle of Willis as shown in Fig. 1 (136,137). Successful perforation

is indicated by a sharp, sudden increase in ICP and an associated decrease in CBF (136,137). Both parameters should both be monitored during the procedure alongside systemic blood pressure, heart rate, and peripheral oxygen saturation (136). After perforation, at which the loss of a slight resistance can be felt, the filament must be withdrawn (136,137). If no increase in ICP is observed, the filament must also be withdrawn, and a second attempt at perforation may be made (136,137). Finally, the filament is completely removed, the ECA is ligated and the skin is sutured (136,137). Moreover, it is necessary to monitor the ICP for at least a further 20 minutes post-operatively to detect possibly secondary bleedings (136,137). After the completion of the SAH perforation procedure, anaesthesia should be terminated using, for example, the selective GABA-A receptor antagonist flumazenil (0.5 mg/kg), the opioid antagonist naloxone (1.2 mg/kg), and the α2-adrenergic receptor antagonist atipamezole (2.5 mg/kg) (136). Furthermore, mice should be extubated upon recovery of motor activity (136).

Figure 1: Schematic illustration of the endovascular perforation SAH mouse model

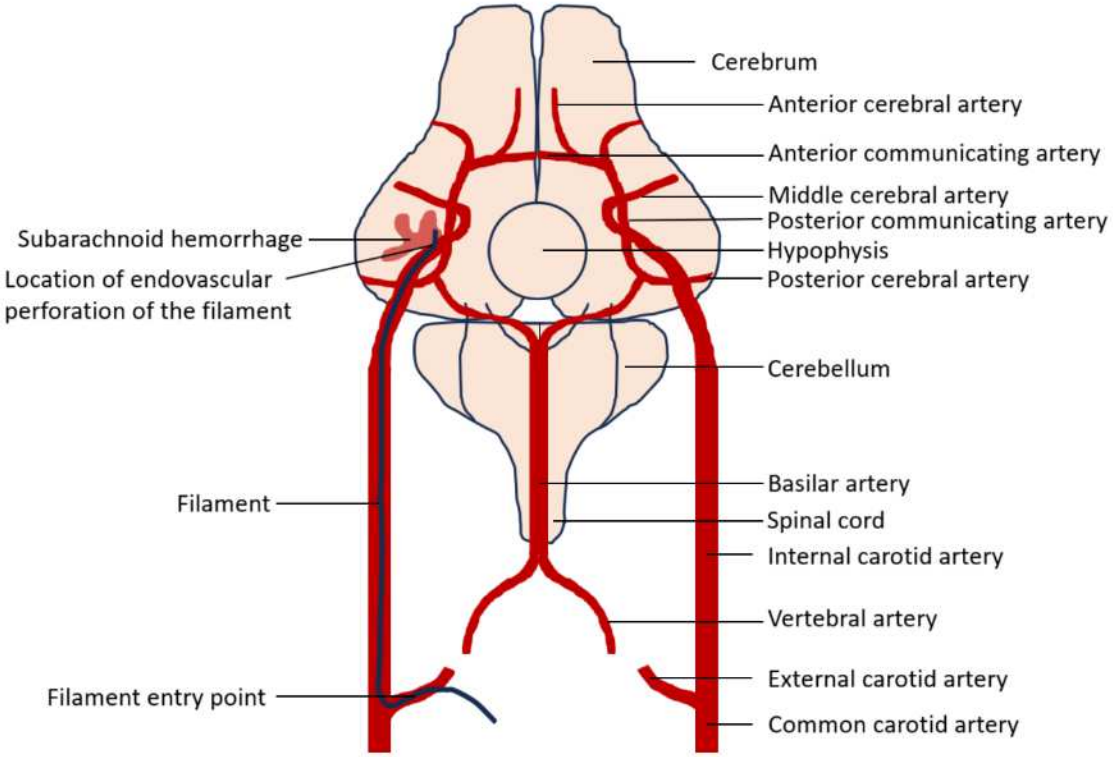


Fig. 1: Schematic illustration of the endovascular perforation SAH mouse model. Own illustration of the procedure and path of the filament according to previous descriptions (44,137). As discussed in this thesis, the localisation of the filament entry point and perforation site may vary (2).

Sham-operated mice, which may serve as a control group for SAH-induced mice, pass all steps of the surgery except the perforation of an artery of the circle of Willis (137). Consequently, the increase in ICP should not occur (137). To ensure sufficient comparability of the groups, it is important to match the experimental animals to the SAH mice in terms of age, weight, sex, and housing conditions (137).

1.2.2.3 Variables within the endovascular perforation SAH model

As the endovascular perforation SAH model is quite complex, it includes several variables that possibly moderate studied outcomes. If these are chosen differently by individual research groups, the reproducibility of the model could be impaired. In the following paragraphs, these potential moderators and the rationale for their inclusion in the review are outlined.

Regarding the experimental animals, besides their species and strain, their sex, age, and weight may influence their health status and thus their condition after SAH induction (136). Underlining the importance of experimental animal selection, a recent study showed sex differences in intracranial aneurysm formation and rupture prevalence (139). Moreover, another recent study highlighted that different strains of mice differ in the anatomy of intracranial arterial vessels which is highly important for the filament path (140). Furthermore, the housing conditions of experimental animals can influence the results of in vivo studies (141). In this context, it is very important whether mice are housed individually or in groups (141). According to the European Union (EU) regulations, "animals, except those which are naturally solitary, shall be socially housed in stable groups of compatible individuals" (142, p. 24). As mice are recognised as social animals, they should be able to engage in species-appropriate social interactions in group housing, which can be beneficial to their health and might improve the transferability of mouse models to humans, who also benefit from social networks (143–145). However, this salutogenic aspect of group housing may be negated by potential aggressive behaviour among animals, leading to increased pain and stress levels (136,146). In any choice, to avoid unnecessary stress around the SAH induction surgery, mice should be housed under constant conditions for at least seven days before and after the SAH perforation (136,146). Another important aspect of animal husbandry is the presence of a circadian rhythm, which can be achieved by automated 12-hour light/dark cycles (147). It has recently been shown that the vascular tone of intracerebral microvessels underlies a circadian rhythm that can modulate SAH damage in the prechiasmatic cistern injection SAH mouse model, where the trauma caused by SAH was greater in the dark phase of the rhythm (147). Moreover, animals should have free access to food and water, as restricted food intake and inadequate hydration can affect the outcome of SAH (148). Blood glucose should also be maintained at a constant level as this can also act as a potential confounder in in vivo SAH models (149). In addition, it has been recommended that mice are placed on a 37 °C plate during surgery to prevent cooling of the animals with temperature monitoring via a rectal probe (136). This may be particularly important for the comparability of results as mice rapidly lose body temperature during anaesthesia (138) and as hypothermia might reduce traumatic brain injury (150). Similarly, differences in the humidity should be considered as a potential confounding factor in the model, as the relative humidity is known to affect the activity and water intake of mice (151,152). Furthermore, the choice of anaesthetic drugs may be relevant for the reproducibility of the model as they have different effect profiles on CBF and therefore might moderate the severity of brain damage caused by SAH (136). For example, isoflurane, fentanyl, midazolam and medetomidine have a particularly low vasoconstrictive effect and are therefore well suited to the model (136). In contrast, propofol and thiopental have a stronger depressant effect on the CBF (153,154). Also relevant is the method used to ventilate the animals during SAH induction as SAH

can cause cerebral ischaemia potentially affecting respiratory brain areas (155). However, there are also variations in the perforation procedure itself that may affect its reproducibility. As there are two locations where the filament perforates vessel walls, namely the entry point into the cerebral circulation and the location of the perforation that is intended to mimic SAH, variations in these perforation sites may lead to different outcomes. In a previous review, besides a considerable share of studies not reporting perforation sites, perforation was often reported at the junction of the ICA with the anterior cerebral artery and the middle cerebral artery (MCA), also known as the *carotid T* (140). However, it is also possible for the filament to slip into either the MCA or ICA and perforate the vessel more distally, resulting in different SAH locations and blood distributions (156). Therefore, objectification of the SAH locations is important, as the low resistance felt by the operator when puncturing the vessel wall, which is often reported in studies as an indicator of correct positioning of the introduced filament and the perforation itself, is not necessarily reliable for a standardised and reproducible induction of the SAH, depending on the operator's experience (140). As a minimum, successful perforation should be monitored with an ICP probe inserted into the epidural space, where, as previously mentioned and illustrated in Fig. 2, bleeding into the subarachnoid space leads to a sudden steep rise in ICP immediately after the perforation accompanied by a reduction of CBF (136). When the filament is retracted, the ICP continues to rise and approaches the level of systolic arterial blood pressure (136). Thereby, the ICP should rise to at least 50 mmHg to presume successful SAH induction (136). Then, the ICP slowly decreases to around 25 mmHg approximately 15 to 20 minutes after perforation (157). After approximately three days, the ICP returns to pre-SAH levels of about 5 mmHg (157).

Figure 2: Intracranial pressure and cerebral blood flow during subarachnoid hemorrhage induction in the endovascular perforation SAH mouse model

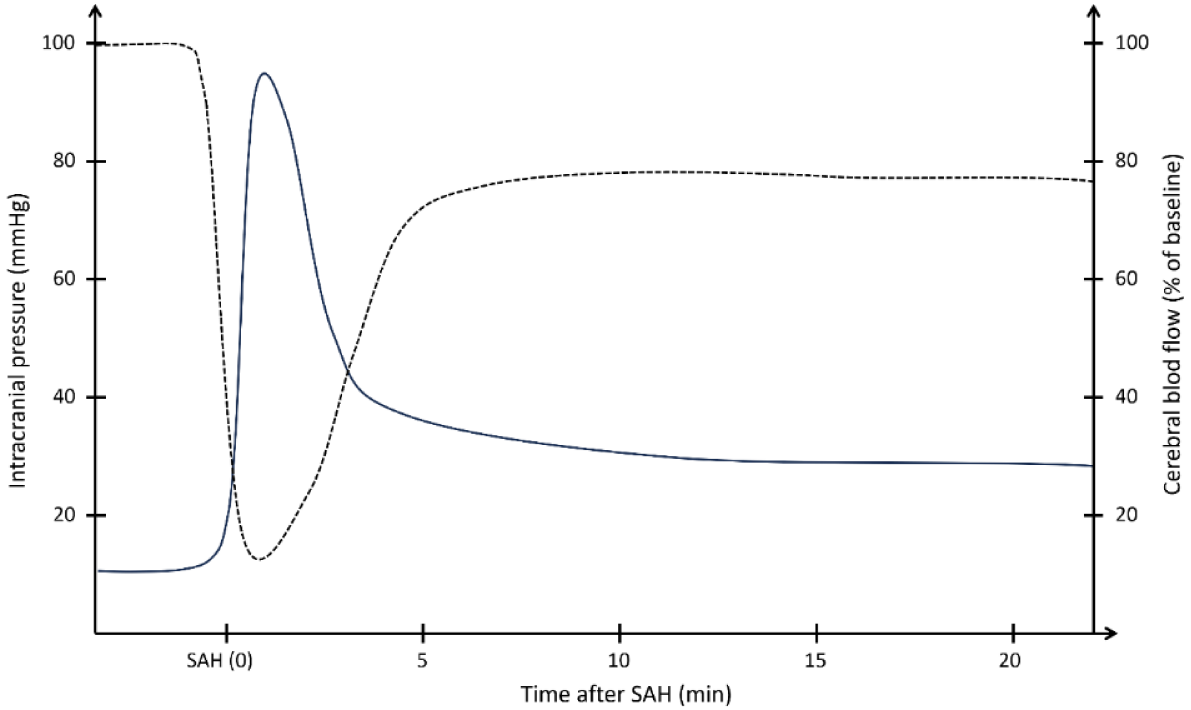


Fig. 2: Intracranial pressure and cerebral blood flow during subarachnoid hemorrhage induction in the endovascular perforation SAH mouse model. The own illustration of exemplary curves shows general trends of the intracranial pressure and cerebral blood flow that have not been experimentally determined during this thesis but are based on external published data (137,158). The dashed curve shows the cerebral blood flow, and the continuous curve shows the intracranial pressure. At time 0, the endovascular perforation is performed and SAH is induced. Cerebral blood flow is expressed as the percentage of the baseline, which is the mean cerebral blood flow before perforation. Min: minutes, mmHg: millimetres of mercury, SAH: subarachnoid hemorrhage.

An alternative way to confirm the successful induction of SAH is to monitor only the CBF without ICP (157). However, cases have been reported where CBF has decreased despite the absence of SAH induction, presumably due to vasoconstriction of the arteries after mechanical irritation where the filament has passed through the lumen (136,157). Therefore, ICP monitoring is recommended (136).

Another factor that could drive irreproducibility across studies is the filament used for perforation as its properties can vary and so may SAH. In particular, this concerns the size of the filament, the material, the tip, and whether it is a monofilament or consists of multiple twisted filaments. For example, a sharpened filament tip may result in a faster healing process of the endothelium after perforation but may also increase the risk of accidental injury to the endothelium when the filament is advanced through the cerebrovascular system (156). Furthermore, a filament with a wider diameter leaves a larger hole in the vessel wall, resulting in increased bleeding into the subarachnoid space and a more demanding healing process (156).

Finally, pain and stress management methods may be another important aspect for the reproducibility of in vivo SAH models, as they cause considerable suffering to the mice (159) and require adequate alleviation (160). For example, depending on the methods used for pain and stress

relief, the cognitive performance of the animals might be impaired to varying degrees, which may confound the results of neurobehavioral tests after SAH (161).

1.2.2.4 Comparative analysis: Cisterna magna blood injection SAH mouse model

As the second main in vivo model type of SAH besides perforation involves injections of blood, this thesis also includes a short review of the cisterna magna injection mouse model of SAH for a comparison with the endovascular perforation model. With this approach, the generalisability of reproducibility findings should be increased and model-specific shortcomings identified.

The cisterna magna injection SAH model models SAH by a direct injection of blood into the cisterna magna (also known as the cisterna cerebellomedullaris) as part of the subarachnoid space while preserving the cerebrovascular integrity (162). The cisterna magna is an extension of the external CSF space and is located dorsal to the medulla oblongata and caudal of the cerebellum, as shown in Fig. 3, and can be punctured percutaneously with the head in a ventral flexed position using a 25- or 27-gauge needle (162). Thereby, the correct positioning of the inserted needle is indicated by the reflux of CSF into the needle (162). Previously obtained autologous blood is then injected into the cisterna magna and the injection is stopped and the needle withdrawn as soon as the ICP reaches the level of the mean arterial blood pressure or the CBF falls below 10 millilitres per 100 grammes of animal weight per minute (162). Afterwards, external pressure is applied to the puncture site and the head is placed in an extended position to allow the injected blood to distribute throughout the subarachnoid space (162). An alternative method of puncturing the cisterna magna is an open surgical approach in which the posterior atlantooccipital membrane is punctured under visual control after skin incision and preparation of neck muscles (162,163). Both procedures should be performed under general anaesthesia (162).

Figure 3: Schematic illustration of the cisterna magna blood injection SAH mouse model

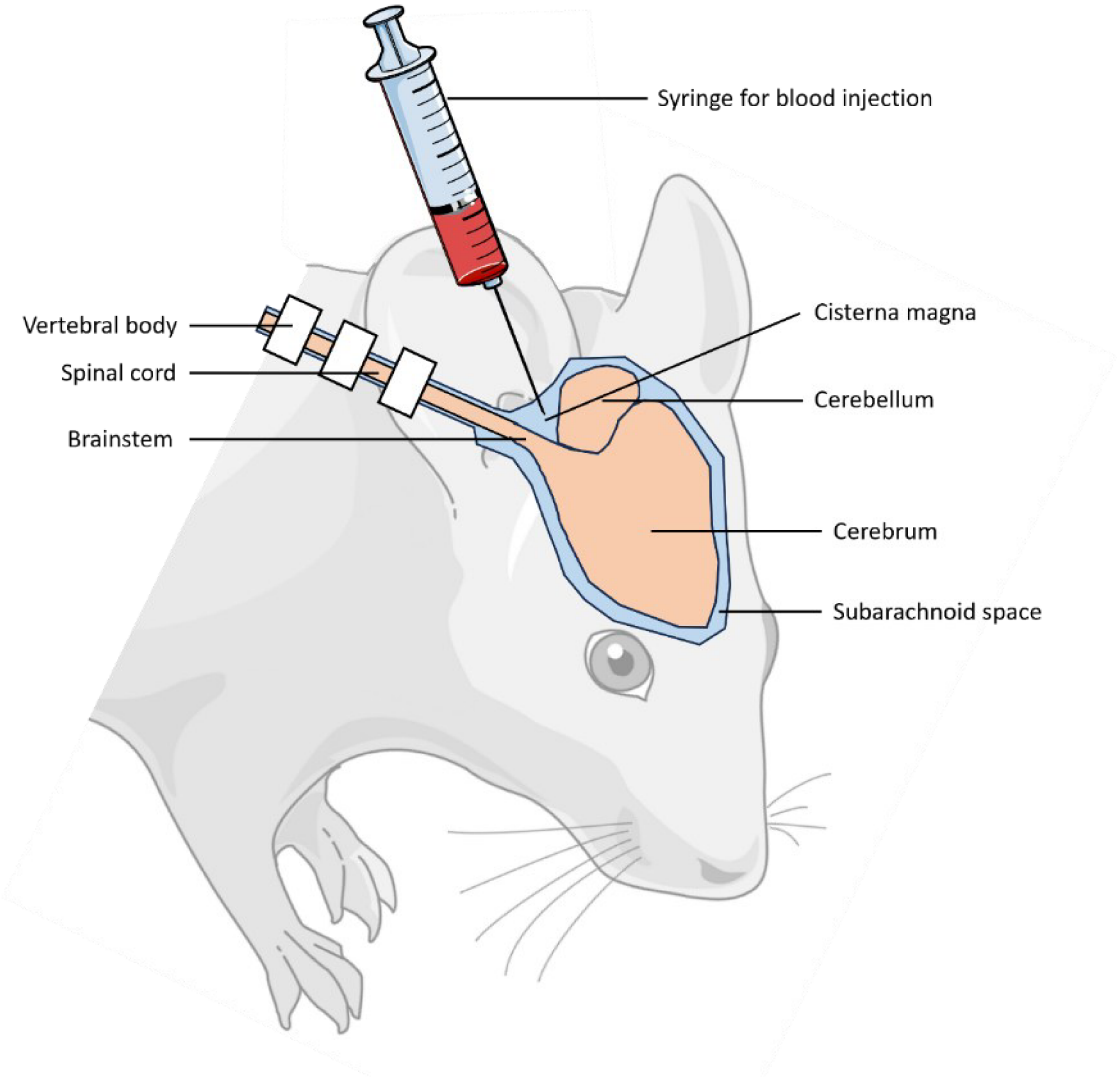


Fig. 3: Schematic illustration of the cisterna magna blood injection SAH mouse model. Own illustration of a simplified procedure according to previous descriptions (44,164). Note that the head is fixed in a ventral flexed position to facilitate access to the cisterna magna. Individual graphical elements were taken from Servier Medical Art (160), which are licensed under a Creative Commons Attribution 3.0 Unported License. SAH: subarachnoid hemorrhage.

Comparing the models, the endovascular perforation model has the advantage of being closer to the natural pathogenesis of SAH and is therefore potentially more representative of the disease in humans (2). However, the extent of SAH is less controllable than in the injection model, as a specified blood volume is introduced into the subarachnoid space, whereas the volume of blood extravasation cannot be directly controlled with endovascular perforation (165). In addition, the perforation model tends to be associated with higher mortality rates than the injection models (2).

1.2.2.5 Investigated outcomes in the in vivo SAH models

For both in vivo SAH models, the outcomes of mortality rate, SAH severity grade, and large artery vasospasm were examined to quantify the reproducibility of reported results across studies (2,3).

1.2.2.5.1 Mortality rate

Animal mortality is a key hard outcome in every in vivo model. Specifically, for SAH models, mortality can be divided into short-term (perioperative) mortality and long-term mortality (in the first weeks after SAH). However, as mice are often sacrificed at the end of experiments to avoid unnecessary suffering in accordance with animal welfare regulations, the observation periods for mortality after SAH induction vary depending on the planned experiments. Determining an appropriate mortality rate is complex. Models with excessive mortality would require more animals for adequate statistical power for hypothesis testing and would also be contrary to the "principles of the 3Rs [sic]" (22, p. 1), that are also incorporated in the EU animal research guidelines (142) and require as few experimental animals as possible (166). However, if the mortality is overly low, this could raise concerns about whether the model is representative of the disease in humans. Overall, assessing the reproducibility of the mortality has the advantage that the outcome is strongly related to the condition of the animals and is commonly reported by researchers, not least because animals excluded from subsequent experiments due to death must be listed in the animal exclusion criteria section for a transparent interpretation of results (40).

1.2.2.5.2 SAH severity grade

Another outcome strongly related to the condition of the animals is the grade of SAH severity (167). Therefore, it was included in the reproducibility analysis to analyse whether SAH could be consistently induced across multiple studies from different research groups. In comparing SAH severities, it is important to consider the various options for its assessment. The induced SAH can be graded by post-mortem histopathological analysis, neurofunctional tests, and various radiological imaging techniques, including non-invasive MRI techniques (168), which allow animals to be kept alive for further experiments (169,170). The probably most commonly used method to assess the severity of induced SAH is the one described by Sugawara and colleagues, in which the distribution of subarachnoid blood is examined via high-resolution imaging, which requires animal sacrifice (167). In this method, the basal cisterns are divided into six segments and the severity of hemorrhage in each segment is defined from grade 0 (no subarachnoid blood) to grade 3 (blood clot occluding all arteries within a segment) (167). The score of all segments is then added together to classify the severity of SAH, as shown in Table 1 (167).

Table 1: SAH severity scoring system according to Sugawara

Evaluation of each segment's score	
0	No blood in the subarachnoid space
1	Minimal blood in the subarachnoid space
2	Moderately severe blood clots with identifiable arteries
3	Blood clots occluding all arteries within the segment
Severity of SAH according to the added score of the six segments	
0 – 7	Mild SAH
8 – 12	Moderate SAH
13 – 18	Severe SAH

Table 1: SAH severity scoring system according to Sugawara. It uses the division of the basal cisterns into six segments and the classification principles according to Sugawara and colleagues (167). SAH: subarachnoid hemorrhage.

1.2.2.5.3 Large artery vasospasm

As mentioned earlier, vasospasm in large cerebral arteries is one of the most dangerous and common complications after SAH (171), is a significant predictor of prognosis, and marks the most important treatable factor contributing to morbidity and mortality in patients with aneurysmal SAH (172). Therefore, adequate modelling of vasospasm in animal models of SAH is highly relevant and thus its reproducibility was analysed in this thesis. As with the SAH severity score, vasospasm can be measured via multiple techniques. It can be measured either post-mortem with the preparation of fixed brain slices by histopathological microscopic measurement of arterial vessel diameters or in vivo with computerised imaging techniques such as MRI and contrast-enhanced angiography combined with digital subtraction angiography (156). In histopathological measurements, perfusion with body temperature-warmed fluid should be used to avoid artificial cold-induced vasoconstriction (173). However, the important disadvantage in comparison to computerised imaging is that multiple measurements on the same animal are not possible due to the need to sacrifice the animal. Moreover, the magnitude of vasospasm can be expressed either as the ratio of arterial diameter before to after SAH or in relation to the arterial diameter in corresponding sham-operated mice (156).

1.3 Aims of the thesis

The primary objective of this thesis was the quantification of the reproducibility of results in representative, commonly used experiments in preclinical neurosurgical research. Irreproducibility was estimated meta-analytically as the variance of published outcomes across studies, beyond what could be statistically expected from random sampling error. The motivation for a precise quantification of reproducibility was to provide concrete data on the prevalence and severity of reproducibility issues, rather than the often criticised "anecdotal" (23, p. 278) nature of irreproducibility reporting so that these can serve as an evidential basis for further discussion and initiatives to facilitate reproducibility in the preclinical neurosurgical science in particular. In addition, in cases of identified irreproducibility, it was intended to identify drivers of limited reproducibility. Thereby, specific model parameter choices in the original studies that lead to reduced reproducibility were of particular interest, so that future targets for

improvements could be highlighted. Furthermore, as a byproduct of extracting data on experimental model characteristics from the primary literature, their frequency of reporting was studied to analyse the extent to which a potentially limited reporting affected the reproducibility of results. Moreover, specific recommendations for the choice of certain model characteristics will be provided if causes of irreproducibility were identified and if a certain choice of a characteristic offered more favourable conditions for the implementation and research on the model, e.g. due to lower mortality of experimental animals. Additionally, common causes of shortcomings in methodological reporting and reproducibility of results in preclinical research will be discussed based on the model reviews' insights to evaluate future perspectives for improving these research quality attributes and subsequently the value of preclinical findings. In total, by investigating and strengthening preclinical reproducibility, the overall aim of this thesis was to contribute to the improvement in the process of translating preclinical scientific knowledge into clinical research and practice to enhance patient outcomes in primarily two very life-threatening diseases, namely glioblastoma and SAH, so that the maximum benefit can be gained from the available financial, animal, and human resources in research (13,15).

2 Meta-analysis on reporting practices as a source of heterogeneity in in vitro cancer research. Sander T, Ghanawi J, Wilson E, Muhammad S, Macleod M, Kahlert DU, BMJ Open Science, Vol. 6(1), pages 1-13, 2022. DOI: 10.1136/bmjos-2021-100272.

3 Meta-review on Perforation Model of Subarachnoid Hemorrhage in Mice: Filament Material as a Possible Moderator of Mortality. Alpdogan, S*, Sander, T*, Zhang, R et al., Translational Stroke Research, Vol. 15, pages 16-29, 2022. DOI: 10.1007/s12975-022-01106-4.

*: These authors contributed equally (shared first authorship).

4 Cisterna Magna Injection Mouse Model of Subarachnoid Hemorrhage (SAH): A Systematic Literature Review of Preclinical SAH Research. Alpdogan S, Li K, Sander T, Cornelius JF, Muhammad S., Journal of Experimental Neurology, Vol. 4(1), pages 11-20, 2023. DOI: 10.33696/Neurol.4.069.

5 Discussion

5.1 Reporting quality

Accurate reporting of the methods and results of studies is a fundamental requirement for reproducible research. However, the reviews conducted within this thesis showed shortcomings in reporting quality for all investigated models (2,3,5).

5.2 Reporting of model characteristics

It is important to ensure that potential variations in the models are accurately reported by the authors, so that other researchers can understand the conditions under which the published results were obtained and reproduce them if necessary.

5.2.1 Reporting of characteristics of the U-87 MG in vitro glioblastoma model

In the articles with U-87 MG cells under TMZ treatment, certain model characteristics were sufficiently reported in the vast majority of included articles, namely the source of the U-87 MG cells, the concentration of TMZ, and the duration of treatment (reported in 94.2 %, 94.2 %, and 87.6 % of the articles) (5). However, there was still margin for improvement as 7.8 % and 12.4 % of the included articles did not mention the concentration and duration of TMZ exposure, respectively, which are both indispensable parameters for the evaluation of treatment response (5). Furthermore, although the source of the U-87 MG cells was frequently reported, 8.0 % of the authors stated that they had received their U-87 MG cells from colleagues (5) which is an inadequate declaration of the source of cells, as it raises concerns about the known problems with genetic alterations, which may increase if the cells are not obtained directly from a certified cell bank (79). Larger deficits in reporting were identified for the cell culture medium and its additives, such as glucose, antibiotics, and FBS (5). In particular, the rather rare reporting of the glucose concentration in the cell culture medium in only 21.2 % of the included studies was of concern (5), as the availability of glucose is important for the energy metabolism of the cells and could therefore influence their response to TMZ treatment (17). Furthermore, shortcomings were identified in the reporting of cell details, such as cell age, cell passaging criteria, cell concentration, and the exclusion of mycoplasma contamination (all reported in less than 20 % of the included articles) (5). In addition, the type of untreated control was not always transparently reported, leaving unclear whether saline, drug vehicle solution, in the case of TMZ usually DMSO, or no additional substances were added to the cells although the choice may influence the relative effect of TMZ on U-87 MG cell viability (5). The infrequent reporting of the added volume of TMZ and control solution (5) might be explainable by the fact that the reported concentrations of TMZ and control solutions could refer to the total suspension of cells, culture medium including additives, and the added drug or control solution, so a statement on the volume would be redundant for the calculation of the actual TMZ and control concentrations in the suspension. However, to avoid misunderstandings, it can be recommended to clarify whether the concentration refers to the combination of culture medium and added drug suspension or only to the added drug suspension itself. Overall, it is evident that there were deficiencies in the reporting of the U-87 MG model that may have limited its interpretability and reproducibility.

5.2.2 Reporting of characteristics of the endovascular perforation SAH model

As in the publications with the U-87 MG model, insufficient reporting was observed for various SAH model parameters (2,5). This concerned, among others, perioperative pain- and stress-reducing procedures mentioned in only approximately a quarter of the studies, and specific information on animal husbandry, temperature, humidity, and access to food and water provided in less than half of the articles (2). Furthermore, at first appearance, most articles reported the location of filament artery perforation and thus the location of the induced SAH (2). However, a closer look revealed that the perforation site was often reported by default at the bifurcation of the ICA and MCA, without mentioning a control procedure to verify the actual perforation site (2). For example, such a verification could be realised by a cranial CT (140). Thus, despite the relatively frequent reporting of the location of perforation, there was a degree of uncertainty as to whether the perforation locations corresponded to those described in the articles. Moreover, as the properties of the filament used for perforation, such as length, diameter, material, stiffness, and tip texture are likely to influence the extent of induced SAH (140), a complete report of these characteristics was expected. However, of those parameters, only the filament diameter was frequently reported in 90.5 % of the articles (2). The filament material and tip texture were reported in approximately half of the reviewed articles, whereas the filament length was reported in only 7.1 % of the articles and no article provided information on the stiffness of the filament (2). In summary, there were also meaningful reporting weaknesses in the literature involving the endovascular perforation SAH mouse model which may have limited its reproducibility.

5.2.3 Reporting of characteristics of the cisterna magna injection SAH model

As only a limited amount of literature was available using the cisterna magna injection SAH mouse model, it was not possible to representatively determine the reporting quality (3). However, observations suggested that overall reporting trends were similar to the endovascular perforation model (2,3). Regarding model-specific parameters, the volume of injected blood was reported in each of the seven relevant articles, whereas the size of the needle used for blood injection was not reported in two studies (3), although it is likely to have an impact on the measured outcomes due to the dependent size of the cerebral trauma.

5.2.4 Reporting of experimental results

In addition to accurate reporting of the model characteristics, a complete and unambiguous presentation of the measured results is similarly important for a meaningful interpretation of study findings.

5.2.4.1 Implementation of independent replications and technical replicates

For every experiment, the number of individual independent experimental units and the number of experimental replications is essential for the informative value of a study, as it usually increases with increasing numbers of experimental units and replications.

In in vivo studies, the number of experimental units (often named *n*) is usually the number of individual experimental animals and has to be reported for each experiment (174). However, if animals

are housed in groups, they may influence each other and thus cannot be considered independent, which may lead to a reduction in the number of experimental units *n* to the number of groups, depending on the magnitude of inter-animal interactions (175). Therefore, in the course of the SAH model reviews in this thesis, the correct number of experimental units was often difficult to assess due to the limited reporting of the presence of group or individual animal husbandry (2,3). Therefore, it is strongly recommended to report the degree of interdependence between animals.

In contrast, in in vitro cell culture experiments, it cannot be reasonably assumed that each cell represents an individual experimental unit *n*, since cell populations contain millions of cells that are studied as a whole, with attention often focused on increases or decreases in the sheer number of cells. In this case, the cell population, mostly cultured in a well or culture dish, can represent one experimental unit *n*. However, it is still important to distinguish between technical and biological replications, as the degree of independence between parallel experiment replications determines their implication for inferential statistics (176). Technical replications occur when replications are not sufficiently independent from each other, for example, when multiple wells on a multi-well-plate are incubated at the exact same time with the exact same reagents simultaneously prepared originating from the same material and cell stock (176). Here, the measured results of the wells can be used to calculate an average to provide a more accurate estimate of the individual replication result and of the variance of the measurement system (176). To narrow the confidence interval (CI) of the overall experimental result, sufficient independence between the replications of the experiment needs to be introduced. This could be realised, for example, by a repetition of the experiment on different days each with freshly prepared reagents, so that each day represents a separate biological replication and counts for the number of independent experimental units *n* (175–177). However, in the review on the U-87 MG model, about two thirds of the articles did not clearly distinguish between technical and biological replications, as almost no measures to ensure sufficient independence were declared (5). In case of insufficient independence between replications that are used for inferential statistics, the phenomenon of "pseudoreplication" (175, p. 1) may occur, where invalid conclusions about the underlying population are drawn based on an overestimated number of individual experimental units *n* (175). In addition, there are different possible ways for calculating the variance of the overall result of an experiment with multiple biological and technical replications, which, if not clarified in the articles, may lead to different estimates of variances and thus may bias conclusions (5). Furthermore, missing or unclear information on the type of result and variance especially in graphical presentations, either as mean or median and standard deviation (SD) or standard error of the mean (SEM), further complicated the interpretation of reported experimental results (2,5). Because the SEM divides the SD by the square root of the underlying sample size of experimental units *n* (178), it is generally smaller than the SD. Thus, error bars without clarification of whether the SD or SEM is presented may lead to biased variance estimates (2,5).

5.2.4.2 Unclear timing of outcome assessments

Another difficulty in the interpretation of published results was the partial absence of reporting the timing of outcome measurements. For example, the duration of TMZ treatment of the U-87 MG cells

prior to viability measurement was not reported in 12.4 % of the reviewed articles (5). Similarly, the time between SAH induction via endovascular perforation and outcome assessments was unclear in 14.3 % of the studies (2). Given that a time dependency of the analysed outcomes was identified in both models, every experiment should include accurate timing information (2,5).

5.2.4.3 Unclear dosage descriptions of tested drugs

Apart from the time at which the results are measured, in intervention studies with administered drugs their dosing is another essential factor for the interpretation of results. However, the review of the U-87 MG model revealed that information on the concentration of TMZ treatment was often missing or incomplete (5). In addition, it was not always clear which specific dosing regimen was used for TMZ application, e.g. single application in the beginning or repeated daily administration (5). Therefore, the calculated dose-response efficacy of TMZ may have been influenced by unreported dosing details that contributed to the heterogeneity across studies.

5.2.5 Heterogeneity of infrequently reported model characteristics

Since it is evident that certain key experimental features were often not reported, it is important to know whether these features were chosen differently in the articles where they were reported because one explanation for the infrequent reporting might have been a uniform selection of a particular experimental feature, so that authors deliberately omitted reporting its choice, assuming that it was clear how the model was constructed due to a common standard regarding this feature in the field. However, the reviews in this thesis showed that it is likely that there were no such standards in the infrequently reported model features as they considerably varied among the articles that did report them (2,3,5). For the U-87 MG in vitro model, this concerned, among other parameters, the cell concentration, which varied by a hundredfold between 5 and 500 cells per microlitre, the cell age that ranged from 3 to 100 passages, and the use of both high (4,500 mg/dl) and low glucose medium concentrations (1,000 mg/dl) (5). For the in vivo SAH models, although rarely reported, mice were kept in both group and individual cages (2,3). Also, the age and weight of the animals varied widely between 7 weeks and 20 grammes to 32 weeks and 35 grammes, respectively (2,3). In addition, the diameter of the filament used in the endovascular perforation model was reported from 4-0 to 6-0 on the US Pharmacopeia scale (179) and composed of both nylon and prolene (brand of synthetic polypropylene) with tips both sharpened and blunted at almost equal frequencies (2). Therefore, it can be assumed that the unreported variation in experimental parameters contributed to the irreproducibility of research results.

5.2.6 Recent improvements in reporting quality for the U-87 MG model

An analysis of the development of reporting quality over the last two decades for literature containing the U-87 MG model showed that improvements were made, but so far only marginal (5). This slight progress may have been the result of initiatives to improve the reporting and reproducibility of preclinical research, but it also highlights their current unsatisfactory implementation. As the number of publications per year of studies using this model has increased over time (5), the lack of meaningful progress cannot be attributed to a lower number of publications in recent years.

5.2.7 Correlation between reporting quality and journal citation frequency

An analysis of the correlation between the reporting quality and the journal impact factor (JIF) of journals publishing articles with the U-87 MG model revealed that the article reporting quality increased with a higher JIF (5). This observation may be partly attributable to more rigorous peer review processes and formal requirements for reporting of methodology and results in higher-impact journals (5). In addition, for the often particularly novel and attention-grabbing findings in frequently cited journals, authors and reviewers may apply higher reporting requirements to increase the trustworthiness of such novel findings. However, it is important to note that the positive correlation, as the improvement over time, was rather small (5).

5.2.8 Reasons for shortcomings in preclinical reporting quality

A major reason for the relatively imprecise reporting of preclinical model details and their results might have been a word limit for publications which is often imposed by journals (180). This requires authors to shorten their publications, often in the methods section, to free up space for a more detailed presentation of their results and discussion, where readers and reviewers are expected to pay more attention. Two examples of journals applying such a limit are the *New England Journal of Medicine* and *The Lancet*, both having a limit of 3,500 words for research articles including the methods section (181,182). Another limiting factor of the reporting quality might have been the lack of reporting standards for preclinical research, particularly for in vitro studies (38,45). This absence may have prevented authors from being sufficiently aware of what aspects of methods and results need to be reported for particular models. Furthermore, for research groups specialising in sub-areas of preclinical research, certain methods may have been taken for granted because of their frequent use and therefore were judged to not require a detailed explanation, so that for readers who do not have this routine, the methods are not comprehensible based on the information in the articles. Additionally, methods explained solely with a reference to previously published studies reduce the accessibility of research if additional access permissions are required for these studies. Another element of infrequent reporting may be the possibility of deliberate non-reporting of certain aspects for two purposes. First, authors could have saved on details as a kind of protection against their methods being reproduced by others, so that their expertise remains exclusive to them. Second, non-reporting may act as a proactive defence against potential attacks on their own published results, which would supposedly be less vulnerable to concrete criticism regarding their methods if reported in less detail. However, it is important to note that these aspects are speculative, with no evidence found for them in the reviews (2,3,5). Overall, the unsatisfactory reporting quality was likely multicausal, with article length constraints and lack of reporting standards probably having been the most important factors.

5.3 Reproducibility of results

As the “methods reproducibility” (27, p. 1) was limited in the representative preclinical models, the following section discusses the reproducibility of results with particular attention to the influence of varying experimental parameters on the irreproducibility across articles using the same models.

5.3.1 Reproducibility of the effect of TMZ on the U-87 MG cell viability

Significant and meaningful heterogeneity across studies was observed for the cell viability reducing effect of TMZ on U-87 MG cells with a 95 % CI from 30.0 % to 37.7 % around a mean reduction in cell viability of 33.8 % compared to the untreated control at the same time of exposure (5). Thereby, I^2 as the indicator of the share of variance attributed to real differences between the experimental setups estimated that 99.5 % of the observed variance was caused by true differences across the studies (5). Even after adjusting for different drug concentrations and treatment durations, the proportion of variance attributable to true between-study differences did not decrease meaningfully (5). This demonstrated that although only experiments with the same basic experiment were included, results varied significantly due to differences between the studies. Thus, the experiment cannot be considered reproducible based on the published TMZ cell viability inhibition results.

5.3.2 Reproducibility of results in the SAH models

Due to the low number of relevant outcome data, reproducibility of results could only be investigated meta-analytically for animal mortality and SAH severity grade in the endovascular perforation SAH model (2). For these two outcomes, significant heterogeneity across studies was observed that exceeded the expected random sampling variation largely (2).

The mean mortality rate after SAH was 21.3 % with a 95 % CI from 17.4 % to 25.8 % and I^2 as the share of variance caused by true differences between the experimental setups of 62.8 %, which was notably lower than in the U-87 MG glioblastoma model (2,5). However, similar to the U-87 MG model, the heterogeneity was highly significant and meaningful ($p < 0.001$) with a 95 % prediction interval for the mortality rate in a hypothetical new experiment ranging from 7.5 % to 47.5 % (2,5). Also, after adjusting for different observation periods between mortality assessment and SAH induction, the review still indicated notable irreproducibility (2). In contrast, mouse mortality in the sham-operated non-SAH control cohorts can be considered reproducible with no statistically significant heterogeneity across articles and a mean mortality rate of 5.1 % with a 95 % CI from 3.3 % to 7.9 % (2). However, it is possible that the exclusion of deceased animals in the sham-operated groups, e.g. due to accidental puncture of a vessel wall, has not been completely reported which has potentially artificially increased the homogeneity in this group. In the literature on the cisterna magna injection model, mortality rates varied between 0.0 % and 22.0 %, but the variance should be interpreted with caution, as three of the four studies with relatively low mortality rates did not explicitly mention an observation period longer than one hour after SAH induction (3). Therefore, it was not possible to draw valid conclusions on the reproducibility of this model, although the descriptive data suggested heterogeneity in this model too (3).

Regarding the SAH severity grading, the system proposed by Sugawara and colleagues (167) was the only one used in a sufficiently large number of studies involving the endovascular perforation SAH model to be included in a meta-analytical evaluation of its reproducibility (2). Thereby, the mean SAH score was 10.7 with a 95 % CI of 9.6 to 11.7 on a scale of 0 to 18, indicating an overall moderate SAH (2,167). In contrast to mortality, the SAH severity score did not show significantly higher variance

across the articles than what could be expected by a random dispersion of results, suggesting that SAH was induced reproducibly in the model, at least on the scale according to Sugawara and colleagues (2,167). However, bias may have occurred by the exclusion of animals with too high or too low SAH scores outside the target range for subsequent experiments. Moreover, animals with a potentially very high SAH score might not have been able to be included in the SAH grading procedure due to premature death shortly after or during the perforation procedure. Furthermore, as usually only a small sample of mice from the total population of a study cohort was used to objectify the severity of SAH, it is possible that animals more likely to suffer moderate SAH were favourably selected for the SAH grading subgroup based on their external appearance, resulting in selection bias. However, none of the studies included in the review of the cisterna magna injection SAH model reported an SAH severity score according to Sugawara (167) or a different frequently chosen grading system, so no statement could be made on its reproducibility (3). The absence of SAH histopathological severity scoring may be explained by a potentially better controlled extent of SAH due to the manual injection of a defined volume of blood, which may have made SAH grading seem less relevant. Nevertheless, SAH grading may still be useful to monitor the actual distribution of blood in the subarachnoid space.

Statements on the reproducibility of the third investigated outcome large artery vasospasm were difficult to make because of the overall limited assessment in the primary literature. Descriptive statistics showed a meaningful variation in both models, with a mean reduction in arterial diameter in SAH mice compared to sham-operated mice ranging from 49.9 % to 90.4 % in the endovascular perforation model and from 47.1 % to 89.3 % in the cisterna magna injection model (2,3). Thereby, observations in the endovascular perforation model indicated that vasospasm decreased with increasing time after a maximum vasoconstriction occurring approximately one day after SAH (2). In addition, the basilar artery tended to have the least vasospasm (2), presumably due to its greater distance from the perforation site. This observation was confirmed in the cisterna magna injection model (3). Likewise, data from one included study suggested that vasospasm was more severe on the ipsilateral side of the perforation (2,183). Due to the central location of the cisterna magna, it was not reasonable to determine the extent of vasospasm depending on the injection site. However, due to the relatively heterogeneous settings in terms of measurement methods, measurement times, and comparative groups, as well as the small animal cohorts, the question regarding the reproducibility of vasospasm could not be sufficiently answered (2,3).

In summary, significant excessive heterogeneity between articles and thus irreproducibility was found only for mortality in the endovascular perforation SAH model, whereas reproducibility could be confirmed only for the SAH severity score in the same model (2).

5.3.3 Identified drivers of irreproducibility in the U-87 MG model

As significant irreproducibility was identified for the cell viability reducing effect of TMZ on U-87 MG cells (5), interest shifted towards drivers of heterogeneity. First, the treatment duration with TMZ was identified as an important determinant of the measured effect of TMZ (5). Thereby, the durations varied widely from 4 hours to up to 12 days with a non-reporting rate of 12.4 % (5). Moreover,

dose dependence was demonstrated and modelled using a four-parameter log-logistic dose-response model (5,184). As these two parameters are well known predictors of efficacy for many drugs, their influence was statistically adjusted in further analyses of reproducibility moderators arising from differences in model characteristics (5). In these analyses, the glucose concentration in the cell culture medium was the only significant moderator of the effect of TMZ and its reproducibility (5). Reflecting different choices of culture medium glucose concentrations increased the explained proportion of true variance across the articles by 3.3 % to 45.4 % (*marginal R2*), indicating its relevance beyond differences in TMZ dose and treatment duration (5). Limiting the explained variance, glucose concentration was reported in only 21.1 % of the 137 articles, with only three articles explicitly mentioning the use of low glucose conditions (1,000 mg/dl), whereas 24 articles described high glucose conditions (4,500 mg/dl) (5). Interestingly, the cell viability reducing effect of TMZ was on average almost 40 % lower in the articles with high glucose concentration than in the articles with an unreported glucose concentration (23.1 % (95 % CI: 15.8 %, 30.5 %) and 37.1 % (95 % CI: 28.6 %, 45.6 %), respectively) (5). The estimate for the effect of TMZ in the articles with low glucose concentrations was not meaningful due to the small number of corresponding articles (5). From a pathophysiological perspective, elevated glucose levels are known to decrease the sensitivity of glioblastoma cells to TMZ, including U-87 MG, as demonstrated in both in vitro (88,89) and in vivo cell transplant conditions (185). In addition, lower blood glucose levels in glioblastoma patients have been observed to be related to higher overall survival rates (186). This phenomenon has been attributed in part to the upregulation of the guanine nucleotide-binding protein coupled formyl peptide receptor one and the epidermal growth factor receptor in the presence of elevated glucose availability, which promotes tumour cell migration and proliferation (88). Moreover, an increased stimulation of the insulin-like growth factor one receptor during hyperglycaemia has been associated with tumour progression (89). Furthermore, the sensitisation of U-87 MG cells to TMZ treatment has been attributed to an enhanced induction of apoptosis due to increased intracellular calcium concentrations in the presence of glucose deprivation (89). Based on the above, it could be hypothesised that the articles not mentioning the glucose concentration predominantly used low glucose concentrations, as the TMZ sensitivity of U-87 MG cells was significantly higher in this group (5). It is possible that low glucose conditions in the cell culture medium were considered by many researchers to be the standard in such cell culture experiments and were therefore not explicitly reported in the publications (5). However, authors who replied to the contact regarding missing information on experimental parameters mainly used high glucose conditions, which contradicts this hypothesis, although the number of responding authors regarding this parameter was rather low with only seven (5).

Overall, it was demonstrated that the glucose concentration in the cell culture medium moderated the response to TMZ treatment, but it remained unclear how the choice of concentration concretely affected the outcome, so that no definite recommendation could be made for the glucose concentration in future experiments with the U-87 MG model, except for an urgent call to report the chosen concentration.

5.3.4 Identified drivers of irreproducibility in the in vivo SAH models

Since significant irreproducibility was only found for animal mortality in the endovascular perforation SAH mouse model, the moderator analysis was limited to this outcome (2,3). As for the in vitro U-87 MG glioblastoma model, a significant relationship between the time of outcome assessment and the value of the outcome was demonstrated (2,5). This was expected as the mortality rate was analysed as cumulative mortality. Concerning model-specific reproducibility moderators, the initial regression analyses did not identify a significant moderator of the variance in mortality across studies (2). The parameters that came closest to significant moderation of mortality were the material and tip texture of the filament used for perforation (p = 0.103 and 0.136, respectively) (2).

As the pathomechanisms leading to death were expected to differ between the very early phase after the onset of SAH and the later phase, a sensitivity analysis was performed by excluding hyperacute mortality data assessed within the first 24 hours after SAH induction (2). With this approach, it could not be strictly distinguished between early and late causes of death after SAH, as mortality data for the first three days after SAH also included earlier causes of death. Instead, an attempt was made to reduce the influence of studies focusing mainly on earlier mechanisms of death, which were more likely related to direct damage of brain tissue by the hemorrhage (187). In addition, earlier causes of death probably stronger involved complications during the SAH induction surgery itself such as thromboembolism, infection, accidental vascular injury, an acute increase in ICP with cerebral herniation, and anaesthesia-related adverse events (187). Thereby, the complication incidence may have been more dependent on the skill and experience of the researchers performing the surgery than on the choice of model characteristics. After excluding mortality data assessed within the first 24 hours after SAH induction and data with unclear observation periods, 34 of 40 studies remained eligible for the sensitivity analysis in which the overall heterogeneity across studies changed only marginally (2). Interestingly, however, a significant relationship was observed between the material of the filament used for perforation and mouse mortality rate, reducing the proportion of remaining true variance due to between-study differences from 59.0 % (95 % CI: 37.1 %, 80.9 %) to 53.1 % (95 % CI: 25.1 %, 77.2 %) of the total variance (2). Thereby, the mortality was highest in the articles with an unreported filament material with a mean mortality rate of 34.4 % (95 % CI: 24.3 %, 44.0 %) (2). For the two reported materials, nylon and prolene, mortality was substantially lower at 19.2 % (95 % CI: 15.0 %, 24.3 %) and 25.0 % (95 % CI: 15.8 %, 37.2 %), respectively, but there was no significant difference between both materials (2). As the data did not provide an apparent explanation for the higher mortality in the subgroup of studies with unreported filament materials, several potential explanations for this observation must be considered. First, there may have actually been a real difference in the SAH-inducing properties and subsequently mortality rates between nylon and prolene filaments that may not have been detectable in this review due to the small number of five articles using prolene filaments (2). Second, studies in the unreported filament material subgroup may have used filaments composed of other materials such as vicryl (brand of synthetic polyglactin 910), monocryl (brand of synthetic poliglecaprone 25), or silk (natural protein fibre), which could potentially have resulted in more severe SAH and higher mortality

rates. In contrast, the use of tungsten filaments and hollow tubes has been described to reduce the risk of inadvertent vessel damage during SAH perforation surgery (188), although none of the included studies reported their use (2). Moreover, it is possible that other currently unknown confounders contributed to the excessive mortality in the unreported filament group or that longer observation periods in this subgroup biased the analysis (2). Among the reported filament materials, nylon and prolene are both non-absorbable and typically of a monofilamentary texture (189). However, prolene filaments usually have a greater tensile strength (189), which may result in greater endothelial damage during perforation of an arterial wall. For nylon filaments, polyfilaments are also available (190), which also tend to cause greater endothelial damage and therefore may increase bleeding severity, potentially contributing to the heterogeneity of model results (2). Overall, there was an indication that the material of the filament may have had an impact on animal mortality and thus, depending on its choice, the reproducibility of the model, although no statements could be made about the concrete influence of specific materials on the extent of SAH.

5.3.5 Additional sources of irreproducibility

As the identified moderators explained only a small proportion of the observed irreproducibility, there had to be additional unidentified factors that led to such heterogeneous results across studies. Failure to identify these may have been due to non-inclusion in the review, inherent heterogeneity of the modelled diseases, and insufficient statistical power due to infrequent reporting in the underlying primary literature. Their potential impact on the reproducibility of the analysed models is discussed in the following.

5.3.5.1 Impact of non-significant reproducibility drivers in the U-87 MG model

First, there is the possibility that the examined model parameters, which theoretically should influence the properties of the models, may still be relevant to their reproducibility, even though no significant moderation of the results was identified in the meta-reviews. Thereby, a main reason for the lack of significance may have been the partly low number of studies in the model parameter specification subgroups. To ensure that their relevance for the reproducibility of the models is not underestimated, they should be examined more closely in future analyses, although this first requires improved reporting of those in the original literature to increase the statistical power. These future analyses may initially focus on those parameters that have been shown to be close to the threshold of significant reproducibility moderation in the analyses conducted as part of this thesis. For the U-87 MG model, this concerns the source of FBS and the U-87 MG cells, as well as the addition of antibiotics to the cell culture medium ($p = 0.067$, 0.075, and 0.094, respectively) (5). As mentioned earlier, the source of FBS is an important factor in cell cultures as the exact composition of its ingredients such as growth factors, hormones and nutrients has been found to vary which possibly leads to different cell proliferation properties (191). Therefore, it is recommended to obtain FBS only from certified distributors and to indicate the source of supply (47,92). Likewise, it is important to obtain U-87 MG cells only from reputable distributors, as variations in the genetic identity of this commercial glioblastoma cell line have been shown (62). Another potential contributor to heterogeneous results may have been added antibiotics, which are used

to prevent bacterial contamination of cell cultures. However, it has been recommended that antibiotics should not be used routinely because of possible changes in gene expression (192) and differentiation of cancer cells (90).

For the endovascular perforation SAH model, filament texture and type of anaesthesia ($p = 0.136$ and 0.163, respectively) were the closest to significantly moderating mortality in mice (2). As previously discussed, monofilaments and polyfilaments may have different effects on SAH because the twisted polyfilaments might penetrate the vessel wall more traumatically, potentially leading to more severe bleeding (156). Regarding the choice of anaesthesia technique in mice during SAH induction surgery, inhalational anaesthesia with isoflurane has the advantage of a relatively faster postoperative recovery and generally lower anaesthesia-related mortality (193). In contrast, for intraperitoneal injection anaesthesia, as chosen by most authors in the review (2), the respective side-effect profiles of the anaesthetics administered have to be taken into account, especially regarding their effect on CBF (136). For example, propofol and thiopental have a relatively strong circulatory and CBF depressant effect (153,154), whereas fentanyl, medetomidine, and midazolam can be rapidly antagonised to terminate anaesthesia and have less cardiovascular and CBF impact (136). Moreover, in comparison to human patients, anaesthetics are also important in the management of SAH to attenuate hypertensive episodes caused by the Cushing's reflex and to reduce ICP to maintain adequate cerebral perfusion pressure and hence CBF (194). It is important to note, however, that the autoregulatory mechanisms of the cerebral vasculature may be impaired in SAH, thus potentially altering the response to anaesthetic agents (194).

5.3.5.2 Inherent heterogeneity of modelled diseases

Besides the characteristics of the models, the inherent biological heterogeneity of the modelled disease themselves may also have led to some degree of irreproducibility (5,195). Thus, it is possible that a preclinical model may not be reproducible per se if the modelled disease is highly heterogeneous. The reproducibility of the model would then only be achievable if the disease is greatly simplified in the model with trade-offs regarding the complexity and heterogeneity of the disease. However, if the aim is to represent the disease precisely, it may be necessary to accept an appropriate degree of model irreproducibility, even if this means more difficult conditions for studying the disease. Indeed, both diseases investigated in this thesis are characterised by a large variance in humans (196,197), which may partly reflect the observed heterogeneity of the models. In glioblastoma, heterogeneity is a key pathophysiological feature, particularly in the development of therapy resistances (60,198). Thereby, heterogeneity refers to the variation of tumours between patients as well as within a single tumour in one patient (60,198). Thus, modelling the heterogeneity of glioblastoma is a critical aspect of preclinical brain tumour research to fully understand the disease. For this purpose, cell cultures derived directly from patient samples have been increasingly used to allow more individualised studies (94). However, it is notable that patient-derived models have not shown a higher variance in TMZ sensitivity compared to conventional commercial glioblastoma cell lines such as U-87 MG (87). Similarly, SAH, although predominantly classified as traumatic or spontaneous SAH and further subdivided into aneurysmal and

non-aneurysmal types, is much more heterogeneous in various aspects including pathogenesis, clinical presentation, treatment, and prognosis (197,199). The extent to which the observed irreproducibility is due to the inherent heterogeneity of the diseases, and the extent to which the variance between studies was due to actual differences in experimental design and procedures, could not be conclusively determined based on the meta-analyses in this thesis. For the estimation of inherent disease-related heterogeneity of a preclinical model, multiple experimental replications with exactly the same model specifications under the same circumstances would be most appropriate.

5.3.5.3 Additional potential irreproducibility drivers that are difficult to capture

It is likely that, in addition to the parameters investigated in the reviews, there were additional drivers of irreproducibility whose influence on the preclinical models could not be adequately determined. One such factor may have been differences in the experience and skill of the experimenters, especially in highly technical models such as the endovascular perforation SAH mouse model, requiring a feeling for the resistance when puncturing the intracerebral artery (140). Moreover, less experienced researchers may have needed more time to complete the SAH induction surgery, which may have resulted in additional stress to the animals due to the longer duration of anaesthesia. Another factor that is hard to account for in reproducibility investigations was the uncertainty of errors in manual outcome assessment procedures, such as manual cell counting or SAH severity score assessment, which may have contributed to the excessive variance in reported results. In addition, minor deviations from the actual experimental protocols may have occurred without full documentation in the primary literature, either because they were not consciously perceived in routine procedures or because they were considered too burdensome for accurate documentation. Furthermore, authors may have been discouraged from disclosing protocol deviations by fears of a perceived devaluation of the quality of their research. Such potential undocumented deviations from the stated methodology could have biased the results of the reviews.

In all these aspects, it is important to note that the difficulty in analysing the impact on reproducibility is mainly a result of the absence of reporting them in the primary literature. This means, on the one hand, there is no evidence of their existence and, on the other hand, that the transparency of scientific publications needs to be improved to evaluate these potential additional drivers of irreproducibility.

5.4 Comparison of findings with external observations

In order to better contextualise the findings of the reviews as part of this thesis, additional external data on the investigated models are presented below which, although not or only partially analysing reproducibility, still contain useful data on the reporting quality and heterogeneity of model results.

5.4.1 Suitable reviews for the comparison

In the following, the studies used for comparison are presented along with their methodological differences from the reviews in this thesis.

5.4.1.1 U-87 MG glioblastoma model

For the comparison of reproducibility data for the U-87 MG in vitro glioblastoma model, only one suitable review was identified with the systematic review by Poon and colleagues, which included 212 articles describing the TMZ sensitivity of various glioblastoma cell lines, including U-87 MG (87). The inclusion of a variety of glioblastoma cell lines marked a major difference from the review in this thesis, with about a quarter of the studies using newer patient-derived cancer stem cells (87). Moreover, only the variance of the half-maximal inhibitory concentration (IC50) data after 24, 48 and 72 hours of treatment was reported and without a reference to an untreated control (87), potentially resulting in a loss of information compared to the data in this thesis in which all TMZ sensitivity data were initially included regardless of the treatment duration and drug concentration (5). In addition, this review did not include a weighted meta-analysis of the extracted data for the analysis of the relationship between different model characteristics and the reproducibility of TMZ response (87). Moreover, neither the development of reporting quality over time nor its correlation with the citation frequency of the publishing journals was investigated (87). Furthermore, some of the important experimental parameters examined in this thesis, such as the glucose concentration, cell line authentication, mycoplasma contamination exclusion, antibiotics supplementation, and cell passage criteria, were not evaluated in the descriptive review by Poon and colleagues (87). In addition, there was a relatively older work reviewing the influence of publication bias on the reported effect of TMZ in in vivo glioblastoma models, which included sixty articles with a total of 2,443 mice and rats (200). In these, U-87 MG cells and other glioblastoma cell lines were transplanted into the animals (200). Although this review did not provide data on in vitro cell culture experiments, the results on the relationship between the methodological quality of the publications and the reported effect sizes were similar to the observations in the review of this thesis (5,200).

5.4.1.2 SAH in vivo models

Three different reviews were identified for the comparison of the reporting quality and reproducibility of the investigated in vivo SAH models. All three reviews were not model-specific but partially contained separate data for each SAH model. First, a systematic review of SAH in vivo models included a total of 765 studies published between 2000 and 2015 using a variety of experimental animal species, including mice, rats, rabbits, and even non-human primates (45). However, this review did not include a meta-analysis of the heterogeneity of results, and the experimental parameters examined differed substantially from those evaluated in this thesis (45). Second, another systematic review investigated the SAH complication DCI in animal models, including a total of 78 studies, of which 39 used an injection model and 23 used the endovascular perforation model in mice and rats (44). However, as in the previously mentioned review, no meta-analysis assessing the reproducibility was performed (44). Moreover, the severity grade of SAH was not considered (44). Third, a review of 48 studies with different SAH in vivo models published until 2015 provided a meta-analysis of mortality variation, but did not include such an analysis for the SAH severity grade and large artery vasospasm (201). Additionally, some experimental parameters that were included in the review of this thesis, such as the

animal housing conditions, anaesthesia, location of the perforation, and size of the needle used for blood injection, were not recorded (201).

5.4.2 Reporting quality

Overall, external reviews confirmed that the reporting on multiple elementary features of the models was inadequate, although there were notable differences in the reporting frequencies.

For the U-87 MG glioblastoma model, one large difference in reporting between this thesis and the review by Poon and colleagues was the reporting of the cell concentration with 72.6 % reporting in the investigated articles (87) compared to 16.8 % reporting in this thesis (5). This discrepancy may have been due to different definitions of sufficient reporting of cell concentration, as in the review by Poon and colleagues, reporting the number of cells per well might have been sufficient to classify the cell concentration as reported, although this was not explicitly stated (87). In contrast, in this thesis, in addition to the number of cells per well, the corresponding well volume had to be reported for the classification as reported, as this is required for the calculation of the concentration, even if a standard well volume might have been assumed by some authors (5). Model parameters with a similar reporting frequency were the clear description of the number of independent replications and technical replications in the range from 20 % to 35 % reporting and both the TMZ concentration and treatment duration in the range of 80 % to 95 % (5,87). Moreover, the reporting of cell passage criteria was only slightly better in this thesis with 12.4 % compared to 7.5 % in the external review (5,87).

Regarding SAH in vivo model parameter reporting, in the review by Grüter and colleagues (45), anaesthetics used during SAH surgery were similarly often reported in around 90 % of the respective articles (2,3,45). In contrast, the age and weight of experimental animals were reported more frequently in 93.9 % of the articles in the review by Grüter (45), compared with 76.2 % reporting animal weight and 66.7 % reporting animal age in this thesis (2). This difference may have been due to the aggregation of animal weight and age into one item, where the reporting of one of these two parameters may have been sufficient for the classification as reported for the whole age and weight item (45). Furthermore, in the review by Goursaud and colleagues, intraoperative monitoring of ICP was slightly more frequently reported in 38.5 % of the articles (44) than in this thesis for the endovascular perforation and cisterna magna injection model with 26.2 % and 28.6 %, respectively (2,3). Interestingly, there was a large difference in the reporting of filament size in the perforation model, as the review by Grüter identified a reporting rate of 43.4 % of the 173 reviewed studies (45), whereas a reporting rate of 90.5 % was observed for the filament diameter in this thesis (2). Again, the reason for the large discrepancy may have been the definition for the parameter classified as reported, as the meaning of size in terms of thickness or length was not specified in the external review (45). However, the range of filament diameters from 4-0 to 6-0 was confirmed (2,45).

Regarding the cisterna magna injection SAH model, Grüter and colleagues discovered that only one third of the studies reported the injected blood volume in various blood injection SAH models (45), whereas this thesis experienced a reported blood injection volume into the cisterna magna in each of the seven relevant articles (3). However, another review confirmed the observed range of injected blood

volumes from 30 to 400 µl, almost identical to the findings of this thesis (3,201). Interestingly, the review by Grüter including multiple SAH in vivo models showed that the reporting quality was independent of the choice of experimental animals and SAH models, and only a slight improvement in reporting was observed between 2000 and 2015 while it was not significantly associated with the JIF or the number of article citations (45). However, it was observed that articles with a higher reporting quality tended to have smaller effects (5). This negative correlation between reporting quality and effect sizes was also found in a review of drug treatment of vasospasm in in vivo SAH models (202) and in various investigations of ischaemic stroke research (203–205). Moreover, this trend was also observed in the review of the U-87 MG cell culture model as part of this thesis, where the effect of TMZ was significantly inversely correlated with the articles reporting quality (5). Hence, studies with lower methodological reporting quality might be more susceptible to overestimation bias of the studied outcomes (45).

5.4.3 Reproducibility of results

Due to differences in inclusion criteria and analysed parameters, as well as the partly descriptive nature, external data did not allow an exact comparison of reproducibility, but they also indicated meaningful heterogeneity in primary model outcomes.

Regarding the U-87 MG glioblastoma model, the descriptive systematic review by Poon and colleagues revealed substantial heterogeneity in the effect of TMZ on U-87 MG cell viability, significantly exceeding the expected level of random variance with an interquartile range from 34.1 to 650.0 µM around a median IC50 of 230.0 µM after 72 hours of TMZ treatment (87). Thereby, consistent with the findings presented in this thesis, treatment duration was found to be a significant moderator of TMZ sensitivity (5,87). However, in the review by Poon, no additional moderator of U-87 MG TMZ sensitivity was identified (87). Moreover, there were no significant differences in the effect of TMZ between different glioblastoma cell lines, with U-87 MG being the most commonly used (87). In an additional attempt to reduce the heterogeneity, only results measured under normoxic conditions via the 3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide (MTT) assay were included (87) but this reduced the observed variance only slightly (87). Thus, it was suggested that other experimental factors not included in the review, as well as the heterogeneity of the cell models themselves, contributed to the variance, supporting the observation in this thesis where most of the variance across the articles could not be statistically explained (5,87). Another relatively older review analysing the influence of publication bias on the reported effect of TMZ in in vivo transplant glioblastoma models found that TMZ significantly prolonged animal survival compared with untreated controls by a factor of 1.88 (95 % CI: 1.74, 2.03) and decreased tumour volume by 50.4% (95 % CI: 41.8 %, 58.9 %) (200). Interestingly, a significant proportion of the variance in results between articles could be explained by differences in study quality characteristics such as randomisation, blinding and sample size calculation (200). Here, again, the difference in the establishment of these standards between in vivo and in vitro experiments was apparent (5,200).

Regarding the SAH in vivo models, similar to the reviews conducted in the course of this thesis, the review by Kamp and colleagues identified a large variation in animal mortality from no deceased animals up to 100 % mortality in individual studies (201). However, with a SD of 2 % around a mean mortality of 21 %, the variance was significantly lower than in the SAH model reviews presented in this thesis (2,3,201). Thereby, mortality rates differed between SAH models, with the endovascular perforation model having a much higher mortality with 24.3 % than the prechiasmatic and cisterna magna injection models with 10.1 % and 3.8 %, respectively (201). As previously mentioned, time after SAH induction was a significant moderator of mortality (2,201). Moreover, confirming the observations of this thesis, neither the strain or sex of the animals nor the type of anaesthesia significantly moderated animal mortality (201). Furthermore, large variation in animal mortality rates from 6.0 % to 65.0 % was found in another review of in vivo SAH-related DCI modelling, with a comparable range of mortality assessment periods (44). However, there has been no meta-analytic evaluation of the magnitude and moderators of mortality reproducibility, and comparisons of large artery vasospasm data were inconclusive due to different measurement methods (44).

5.5 Challenges in the translation of preclinical findings into clinical practice

The low translation rate of promising preclinical findings into clinical application in human patients has mainly been attributed to the limited reproducibility of research findings (84,206). Given the previously mentioned slow advances in the treatment of the very fatal diseases glioblastoma and SAH, the question arises as to what factors other than the limited reproducibility may be hindering the translation of preclinical findings and how they might be optimised for future improvements.

5.5.1 Disease modelling quality

One key aspect of model-based preclinical research is the modelling quality of the disease being studied in the model. However, it is in the nature of models that they cannot fully capture every aspect of a disease that occurs naturally in humans. Therefore, when testing hypotheses or developing new diagnostics and therapeutics, it is important to use different models to balance their respective weaknesses and strengths (30).

5.5.1.1 U-87 MG glioblastoma model

Regarding the U-87 MG glioblastoma model, the aforementioned heterogeneity of the disease, as well as the observed variations in the genetic identity of the commercial cell line itself are major concerns for the modelling quality (60). Another aspect is the limited ability to model the tumour microenvironment, which is composed of tumoural and non-tumoural cells such as neurons, astrocytes, microglia, and connective tissue cells, and is essential for the formation of a protective milieu and the invasive properties of the malignancy (94,207,208). In addition, oxygen and nutrient availability, which is often reduced in the human tumour environment, has often been kept constant in preclinical models, which may also affect their translatability, as cellular properties are able to adapt to different environmental conditions (5,207). Moreover, modelling the blood-brain barrier is a challenge, especially in in vitro brain tumour research (207). A general issue with the long-term use of conventional cell lines

is the need to immortalise the cells, which can be achieved, for example, by transfecting an oncogenic viral gene (30). However, such genetic modification may also alter the properties of the tumour cell (30). To address the shortcomings of the still widely used conventional commercial glioblastoma cell lines such as U-87 MG and U-251 MG, the focus of brain tumour research has increasingly shifted towards newer in vitro models with a potentially better representation of the disease (60). Such newer approaches include patient-derived cell models, in which tumour material is taken from patients with glioblastoma and is processed for cell culture and in vivo transplantation experiments (93). These cells are then cultured under serum-free conditions mainly in a neurobasal medium containing DMEM and Ham's F12 that promote the development of stem cell properties, inhibit dedifferentiation, and allow the formation of three dimensional neurospheres with the advantage of possible migration process studies (93). It has therefore been suggested that they are more representative of glioblastoma and better reflect tumour heterogeneity (94,207). Thus, although an external review did not observe larger heterogeneity in TMZ sensitivity for patient-derived cell lines (87), it is likely that accurate reporting of model parameters may become even more important for the reproducibility because of the increased tumour heterogeneity (94,207). However, conducting a meta-review for the newer stem cell-like patient-derived cell models, analogous to the U-87 MG review, did not appear reasonable in the course of this thesis due to the large diversity of models and the overall small number of studies per model, resulting in an estimated too low statistical power (5).

5.5.1.2 SAH in vivo models

For the translatability of in vivo SAH model findings to the disease in humans, particular attention should be paid to the characteristics of the experimental animals, differences in their cerebrovascular anatomy and SAH location, as well as the timing of outcome assessment. Regarding experimental animal characteristics, the selection of almost exclusively male experimental animals (2,3) contradicts the clinical reality, in which the majority of SAH patients are female (59,131,209,210). A primary reason for not selecting female experimental animals is the interference with the female hormonal cycle (59). However, the relevance of the hormonal status in humans is suggested by the observation that women are less prone to stroke before than after menopause (59,211). Therefore, research on female experimental animals should be promoted, especially because of the much smaller amount of currently available evidence. Moreover, there are also considerable differences in the ages of laboratory animals and human SAH epidemiology. Mice were used at the beginning of their lifespan (2,3), whereas the median age of SAH in humans has been reported at around 50 to 60 years (103,106–108). This is particularly important with regard to the high incidence of cardiovascular comorbidities in SAH patients, such as arterial hypertension, dyslipidaemia, diabetes mellitus, chronic heart disease, and obesity, which are important in the pathogenesis of SAH and are more frequent with increasing ages (212). Therefore, the use of young healthy mice may have impaired the clinical translatability (211,212), especially as it has already been shown that these comorbidities also influence the development of SAH in animal models (212). However, older experimental animals would be associated with higher husbandry costs and an increased likelihood of external influences on animal characteristics. Hence,

each study should consider the extent to which the age of experimental animals might influence the study objectives individually. In addition, differences in cerebrovascular anatomy between experimental animals and humans in in vivo models of SAH may be relevant for the translatability. For example, the posterior cerebral artery in mice does not arise directly from the BA as in humans, but has been reported to be supplied indirectly via the superior cerebellar artery and the posterior communicating artery (213,214). The extent to which this affects the modelling quality of the endovascular perforation model has not yet been investigated. However, as the posterior cerebellar artery lies not directly in the path of the advanced filament, this deviation may not be of direct relevance, but could result in different ischaemic characteristics due to the different blood circulation. Furthermore, an external review described that nearly 60 % of SAH animal models involve an SAH induction in the posterior part of cerebral circulation, whereas, in humans, SAH location has been reported predominantly in the anterior part of circulation in approximately 90 % of cases (44). However, in the review of the endovascular perforation SAH mouse model presented in this thesis, most perforations were reported in the ICA and MCA bifurcation region which belongs to the anterior part of circulation (2), in contrast to the cisterna magna injection SAH model where the SAH induction occurred in the posterior part (3). Given the close relationship between brain areas and their functions, it is conceivable that this discrepancy is another factor potentially limiting the translatability of certain in vivo SAH models.

Considering animal mortality as a hard outcome in in vivo and clinical SAH studies, it was significantly lower in the investigated mouse models than in humans, by up to a factor of two (2,3,103,129–131,201). This discrepancy may be explained by safeguards to reduce animal stress and suffering during studies, as well as the overall goal of a low mortality in animal studies to reduce the number of animals needed for hypothesis testing (201). However, it is currently unclear whether and to what extent the mortality difference limits the comparability of SAH between laboratory mice and humans (201). Nevertheless, one important aspect partially explaining higher mortality in humans are the longer observation periods after SAH, which are usually much longer in humans than in the experimental laboratory setting (2,3,201). Furthermore, the definition of DCI as a major determinant of SAH morbidity and mortality (117) differs between both fields. In studies with mice, DCI is usually defined as the occurrence of changes in the animal's neurological status with a temporal distance of at least six hours after SAH (201), whereas in humans, DCI is defined as the occurrence of new neurological deficits lasting at least one hour and a radiologically proven new ischaemia or infarction in the original SAH area within the first six weeks and at least 48 hours after SAH (44). The reason for cut-off after six hours is that the mechanisms leading to DCI in humans, which are predominantly vasospasms (140), have been shown to occur earlier in mice with a peak at around eight days after SAH (44,201,215). However, overall, vasospasm appears to be less severe in mice than in humans (140). Hence, differences in the manifestation of DCI as well as too short observation periods in in vivo studies may have contributed to difficulties in the translation into the clinic.

5.5.2 Clinically relevant drug concentrations in the U-87 MG glioblastoma model

Important features for the translation of preclinical drug testing studies are the chosen drug concentrations and the duration of drug exposure for the assessment of a dose-response relationship. However, in the U-87 MG glioblastoma model, the researchers' choice of these two parameters for TMZ treatment may have limited the translatability (5). Here, the vast majority of published studies used much higher TMZ concentrations, namely a median of 100 µM (5), than those achieved in tumour tissue and CSF in human glioblastoma patients treated with TMZ, namely 10 to 50 µM (216,217). These higher doses of TMZ in the in vitro setting have also been observed in other reviews (87,218), although the use of realistic TMZ concentrations close to those achieved in humans has been recommended to maintain comparability with the clinical setting (5,219). In addition, there was concern that the drug carrier solutions might interfere with clinical applicability, as DMSO, the carrier solution for TMZ, has been shown to have cytotoxic effects (219). To counteract this confounding factor, the use of lower physiological concentrations of TMZ may have the additional advantage that water could be used as the carrier solution (219).

5.5.3 Appropriate investigation periods in the U-87 MG glioblastoma model

Theoretically, the full effect of TMZ as an alkylating agent should be expected after the deoxyribonucleic acid replication of all tumour cells has taken place, which can be expected after 34 hours, as this is the approximate cell cycle duration for U-87 MG cells (5,220,221). In humans, the treatment of as many tumour cells as possible is achieved by a continuous administration of TMZ over multiple weeks and an optional simultaneous radiotherapy (222). However, in U-87 MG cell culture experiments, a sufficient incubation time of the cells with TMZ was not always achieved, so that in some experiments the longest treatment time before cell viability measurement was shorter than the assumed 34 hours needed for the full effect of TMZ (5). These short observation periods of sometimes less than 24 hours of TMZ exposure were also observed in an external review (87). As a result, it could be assumed that often only partial effects of TMZ were measured, because TMZ could not develop its full effect due to asynchronous cell cycles within a tumour. Thus, particularly in comparative trials of new drug candidates with a potentially earlier onset of action than TMZ, too short treatment durations may falsely favour new drug candidates over the standard alkylating agent TMZ and thus may impair the translatability of in vitro findings to subsequent clinical studies (5).

5.6 Limitations of the reviews

Regarding the informative value of this work on the reporting quality and reproducibility of preclinical research in neurosurgery, certain limitations need to be mentioned, which result from both the underlying primary literature as well as the applied methodology in the reviews themselves.

5.6.1 Selection of systematic review in- and exclusion criteria

In the evaluation of inclusion and exclusion criteria for the systematic literature search, a balance had to be found between ensuring a necessary degree of homogeneity of the included studies to justify an aggregated meta-analysis and avoiding an unjustified exclusion of evidence relevant to the

objectives of the review by applying too narrow criteria. In the following, certain criteria choices with regard to potential limitations of the reviews are discussed.

5.6.1.1 Potentially too narrow criteria

For the in vitro glioblastoma review, one criterion that may have been too narrowly defined was the inclusion of only one cell line with U-87 MG as the most commonly and still widely used conventional glioblastoma in vitro model (5,87). While the focus on a single model may have resulted in a lower generalisability of the review's findings, it seemed reasonable to choose this model as a starting point for reproducibility studies in this field. Furthermore, due to the similarity in the basic structure of many in vitro cell culture models of glioblastoma, it was expected that the basic findings should be transferable to similar models. In addition, the focus on DMEM as the most commonly used cell culture medium (5,87) was intended to avoid confounding by different culture media and thus, as with the exclusive consideration of U-87 MG cells, selected to provide the necessary homogeneity of studies for the meta-analysis. Because of the same rationale, the meta-reviews of in vivo SAH models only considered mice as experimental animals, also potentially limiting the generalisability of the reviews (2,3). However, mice are a popular choice for these types of experiments due to their relatively low husbandry costs and relatively good cerebrovascular comparability to humans, despite the aforementioned differences (223). Overall, besides providing the necessary homogeneity for subsequent analyses, the mentioned restrictions were implemented to achieve an efficient balance between the expected knowledge gain and the resources required to conduct the reviews, especially in a research area with relatively little pre-existing evidence on reproducibility, where the findings of this thesis can be considered as an explorative starting point for further investigations.

Another concern regarding potentially too narrow inclusion criteria for the U-87 MG review was the required presence of an untreated control group with U-87 MG cells incubated without TMZ, which reduced the amount of TMZ sensitivity data in the meta-analysis (5). Although TMZ sensitivities were also reported without an accompanying control, e.g., by comparing the population growth at the time before the start of treatment, these were not included in the review to ensure sufficient comparability of TMZ sensitivity data and since the comparison with an untreated control is the standard in this type of study (5). However, the reference to the control group and the resulting standardisation of TMZ efficacy may have aggravated the identification of reproducibility moderating variables if they have had the same relative effect on the cell growth in the treatment and control groups, so that the relative efficacy could have remained constant when the confounding variable varied. In contrast, for the SAH in vivo reviews, studies presenting outcome data without a sham-operated non-SAH cohort were included in the analyses, as the outcomes investigated were considered to be more independent of the sham-operated group, and the SAH cohort itself already provided an untreated control for comparison with SAH-induced animal cohorts receiving additional treatments (2,3).

5.6.1.2 Potentially too wide criteria

In addition to potentially too narrow criteria, there may also have been overly inclusive criteria that limited the power of the meta-analyses by introducing excessive heterogeneity. One such aspect

may have been the inclusion of different measurement methods for cell viability in the glioblastoma review, as the measured viability has been shown to differ depending on the chosen assessment method (47,224,225). Likewise, the SAH reviews included both post-mortem microscopic histopathological and in vivo brain imaging techniques for the quantification of large artery vasospasm, despite evidence of differences between these methods (2,3,226). As statistical testing for differences between these two methods was not appropriate due to the small number of articles (2,3), the general premise was to make as few a priori exclusions in the outcome assessment methods as possible, as the frequencies of the measurement methods could not be reliably estimated in advance, and thus the risk of too much loss of information was minimised. However, necessary exclusions were made, for example for clonogenic assays in the U-87 MG review (5), because they typically measure later effects after several weeks, and because they were reported to be more suitable for assessing the effects of radiation than of chemotherapy and, overall, are very different in principle from metabolism-based assays such as the MTT (227).

Furthermore, there were no initial restrictions on the included observation periods and, for the U-87 MG model, on TMZ doses, although these were expected and subsequently proven to moderate the investigated results (2,3,5). This moderation affected the estimation of the extent of irreproducibility and its drivers, which could have been countered by restricting to certain commonly used specifications that would have simplified the meta-analytic model. However, this would have resulted in a loss of information, which should be avoided in meta-reviews when addressing this issue (228). Nevertheless, in the endovascular perforation review, a loss of information occurred in the case of multiple outcome assessments and animal cohorts within a study to ensure sufficient independence of the data included in the review (2). The exact procedure, with its advantages and disadvantages, is discussed in a later section of this thesis on the appropriateness of the applied meta-analytic model.

5.6.2 Restrictions regarding reviewed publication types

Ideally, the literature on which a systematic review is based on should include not only peer-reviewed research articles but also the so-called “gray literature” (229, p. 1) covering scientific data published outside of scientific journals (228). The inclusion of these sources is often intended to reduce the risk of selective reporting bias, a phenomenon in which preferential positive study results are more frequently published than negative results, which can lead to a biased higher overall effect estimate (230). However, the meta-reviews in this thesis only considered data from peer-reviewed research articles, as their purpose was to quantify reporting and reproducibility in precisely this highest quality publication form of primary research (2,3,5). In the non-peer-reviewed forms of publication, it could be assumed that the reporting quality tends to be lower due to a lower implementation of reporting and bias reduction standards. Moreover, the heterogeneity might be greater because of the more frequent reporting of smaller effects probably widening the overall range of published results.

5.6.3 Definition of reporting quality

For the quantitative analysis of reporting, it was necessary to define a score that summarises the quality of reporting (5). Thereby, the parameters included in this reporting score could not be selected

based on a commonly established reporting guideline due to its nonexistence. Instead, the included parameters were combined from various guidelines and best practice recommendations, the experimental experience of the supervising research group, and the potential impact on the investigated outcomes based on previous evidence. Of course, the reporting quality scores and their associations with publication year and JIF probably would have been different with alternative reporting quality score definitions. Nonetheless, with the present selection, an attempt was made to reach the optimum compromise, considering previous knowledge of the importance of the model parameters. Moreover, although certain parameters may have had more influence on the model results than others, each was attributed equal weight in the calculation of the score, as no reliable rationale for an a priori weighting was available (5).

5.6.4 Low response rate by original study authors

When authors were asked to provide information on parameters not reported in the original studies to increase the amount of analysable data in the review on the U-87 MG glioblastoma model, the response rate was very low with only eleven of 137 authors responding (5). Due to the small amount of additional information gained compared to the relatively high effort, authors were not contacted during the following SAH reviews (2,3). A possible reason for the low response rate may have been that some articles were published more than a decade ago, so the contact details provided may not be up to date, or researchers may have retired. In addition, some authors may have been overwhelmed with questions about previous experiments, even if they were limited to a few selected elements, or they may have felt no incentive to respond. Moreover, authors may have been concerned about answering questions inaccurately, which could expose potential weaknesses in their studies, even if this was not the intention of the review. Overall, a higher response rate would have increased the statistical power of the meta-analyses and thus the likelihood of identifying reproducibility moderators.

5.6.5 Selection of a suitable measure for reproducibility

To address the questions raised in this thesis, the definition of an appropriate measure for the reproducibility of the models was critical. As mentioned in the introduction, a distinction has been made between various types of reproducibility (27). It could be argued that “inferential reproducibility” (27, p. 1) is the most important type of reproducibility in hypothesis testing studies, as the conclusions about whether the null hypothesis is maintained or rejected are the key findings of any study and should therefore be identical across studies testing similar hypotheses. However, the experiments investigated in this thesis were mostly not the main aspect of the included primary studies, but rather served as a comparison subgroup for newer approaches. In addition, as the results of hypothesis tests are dichotomous without further gradation, a meta-analysis analysing the influence of individual experimental parameters on “inferential reproducibility” (27, p. 1) would not have been reasonable. Although, for example, the p-value for the difference between the comparison groups of U-87 MG cell viability with and without TMZ treatment could have been used as an effect size in a meta-analysis, this was considered inappropriate because of its low informative value regarding the efficacy of TMZ due to its dependence on many variables that are primarily dependent on study design. In contrast, the

quantitative analysis of primary study outcome data allowed a comprehensive comparison across the studies regarding “results reproducibility” (27, p. 1), and as the decision on hypothesis testing is based on the numerical experimental results, "inferential reproducibility” (27, p. 1) was indirectly assessed in the analyses too.

5.6.6 Appropriateness of the meta-analytical model

In addition to an appropriate effect size for assessing reproducibility, the chosen meta-analytic model is equally important for answering the questions of this thesis. For this purpose, a random-effects model was used because, unlike the fixed-effects model, it assumes real differences between the included studies and thus allows for different true effects estimated individually for each included study (231). However, the interdependence of multiple data within the included studies is a common problem in meta-analyses that can lead to biased results with a falsely narrow CI and to low variance for the overall effect if not adequately addressed, either by using multi-level meta-analytic models or by selectively including representative data from the articles (232). Due to the nature of the extracted data, both approaches were used in the reviews. For the U-87 MG glioblastoma model, a three-level random-effects meta-analysis was performed, including multiple data from each study and calculating an estimate for both within-study and across-study variance (5,233). For the endovascular perforation SAH model review, representative data were included according to the criteria described in the review article because experimental cohorts were often not clearly distinguished from each other (2). This approach was chosen to avoid an erroneously high number of included experimental animals in cases where the smaller cohort was a subset of the larger cohort rather than a group of separate additional animals, which would have led to an overestimation of the weight of the study in the meta-analysis. Furthermore, for the same reason, the proposed approach of pooling mortality data from multiple animal cohorts of a study was not followed (228). No meta-analysis was performed for the cisterna magna injection SAH mouse model due to the small number of included studies and outcome data (3). Although these compromises led to some loss of information, this was preferable to the risk of multiple inclusions and thus incorrect assumptions about the conditions of the meta-analysis.

The metrics tau^2 and I^2 served as well-established measures of the variance of results across the included articles and were therefore suitable for the estimation of the irreproducibility of results (234), as they indicate the amount and proportion of heterogeneity across the included studies beyond what could be expected due to random sampling error (231). As there is no universal definition of irreproducibility, it should be noted that these statistical measures are of course only one of many possible approximations. However, they appeared to be particularly appropriate, although they may have been biased to some extent by overly outlying effect sizes in certain studies.

Furthermore, as multiple meta-regressions were performed in parallel on the same dataset when analysing the influence of the experimental parameters on the results and thus their reproducibility, a correction of the significance threshold could have been useful to keep the risk of false-positive results constant. This may have been achieved, for example, by using the more conservative Bonferroni correction or the Benjamini and Hochberg false discovery rate method (235). Applying these correction

techniques, the variables glucose concentration and filament material, which were found to be significant moderators of the respective outcomes (2,5), would have fallen below the adjusted significance thresholds. However, as the reviews were more of an exploratory nature to assess potential drivers of irreproducibility in the preclinical setting, the classification of parameters as not relevant to the reproducibility of the models based on such corrections might have led to the impression that they were irrelevant regarding the reproducibility of the models, although strong evidence for their relevance was present. Moreover, a cost-benefit assessment before the implementation of multiple testing adjustments is recommended and should consider the potential consequences of an increased likelihood of false-positive results versus an increased likelihood of missing potentially significant results (235). In this thesis, the cost of an increased likelihood of false-positive results was rather low, as a misclassification of presumed moderators of reproducibility might lead to sensitisation and follow-up studies on the selection and reporting of these parameters, but as these further investigations are nonetheless necessary, no resources would be wasted. Moreover, the purpose of the reviews was not to question the results of individual studies or to exactly quantify the effect of a new treatment, but rather to find ways to improve the overall processes of preclinical research. However, to reflect this limitation, and because of the potential confounding of different outcome assessment times and the initially unplanned sensitivity analysis in the endovascular perforation SAH model review, no specific recommendations for the choice of specific experimental parameters were made in this thesis, but attention was raised to improve reporting and to reflect the choice of potential moderators in future studies with these models (2,5).

5.6.7 Most of irreproducibility remained unexplained

One of the aims of this thesis was to identify causes for irreproducibility of the model results to derive strategies for future improvements. However, this was only partially successful for both the in vitro glioblastoma and the in vivo SAH models, with 89.1 % and 90.0 % of the heterogeneity across studies remaining statistically unexplained, respectively (2,5). With the potential reasons for this unexplained heterogeneity already discussed in the previous chapters, possible approaches for much needed future reproducibility investigations are highlighted, including prior improvements in reporting of experimental details and understanding the intrinsic heterogeneity of the disease and the model itself.

5.6.8 Inaccurate and incomplete reporting as a limiting factor in meta-reviews

The partly very low reporting prevalence of certain model parameters complicated the analyses of reproducibility and its moderators in several ways, which is a common challenge in meta-reviews in general (2,3,5,87,201). First, as previously mentioned, one consequence was a reduction of statistical power in the meta-regressions aggravating the identification of reproducibility moderating parameters. In addition, apart from identifying an overall moderation of effects, differences between specific parameter subgroups were hard to detect, with sometimes only one or a few authors explicitly reporting a particular parameter choice (2,5). Often, the subgroup of articles with an unreported parameter choice was the largest, which further complicated reproducibility analyses, as this group was included in the regression as homogeneous although it was likely heterogeneous, as it could not be assumed that each

study used identical experimental parameter characteristics. Moreover, inaccurate information in the original articles may have biased the results of the meta-analyses and reduced their power due to necessary exclusions from and compromises during the analyses, which may have been the case in particular for outcome assessment times and TMZ concentrations, but also for model parameters such as cell passage and concentration (5). Furthermore, certain parameters such as animal age and weight were often reported in ranges without a mean (2,3,5), leading to the trade-off of including the midpoint of the ranges in the meta-regressions (2) resulting in an avoidable loss of information. In addition, if the type of error bars around the mean was not specified, the SEM was assumed as the more conservative approach compared to the assumption of an illustrated SD, which, however, resulted in lower weights for the respective studies in the meta-analyses due to higher variances in case the authors actually presented the SD (2,3,5).

5.6.9 Uncertainties regarding the identified reproducibility moderators

Significant reproducibility moderation was observed for both the glucose concentration in the U-87 MG cell culture medium and the filament material used for perforation in the SAH model (2,5). However, uncertainty remained about the impact of different specifications for both of these parameters, as significant differences could not be identified between the different specification choices (2,5). Instead, significant differences in outcome moderation were found for both models between the large group of articles with unreported parameter specifications and one commonly reported specification, namely high glucose conditions and nylon filaments, respectively (2,5). This was probably due to the very small number of articles with the remaining specifications (2,5). Therefore, only indirect conclusions about the influence of each parameter could be drawn, and no evidence-based recommendations for the choice of any particular parameter could be made. Instead, it highlighted the need for researchers to enhance the reporting of these two model parameters in future studies so that their influence can be more accurately analysed.

5.6.10 Deviations from a priori study protocols

Due to the circumstances that arose during the conduct of the reviews, some adjustments had to be made in comparison to the predefined study protocols published before the conduction of the reviews in the *Open Science Framework* (236,237). In the U-87 MG review, contrary to the original design, animal xenograft models were not included in the analysis, as these were considered to be too different from the in vitro cell culture studies during the preliminary literature research to be included in a joint review (5,236). Furthermore, the analysis of the relationship between the JIF, year of publication, and the reporting quality of the articles was not planned in advance but was later judged to be useful for analysing the context and temporal development of reporting (5,236). Moreover, contrary to the protocols, no analysis of possible publication bias was carried out during the reproducibility assessments (236,237), as publication bias are particularly relevant in meta-reviews that aim to precisely quantify the effect of an intervention, which was not the intention of the reviews presented here (2,3,5). In addition, the irreproducibility of the investigated outcomes was already demonstrated (2,5), so that even greater heterogeneity with smaller effects could be expected in the hypothetical presence of publication

bias, so that the additional analysis of publication bias probably would not have diminished the general findings of the reviews. Furthermore, for the in vivo SAH reviews, SAH severity was chosen as the analysed outcome for the extent of induced SAH rather than the previously intended infarct size (2,3,237), as this was more frequently reported in the literature during preliminary searches. It is also important to note that the sensitivity analysis excluding hyperacute mortality data was not planned in advance and should therefore be given less weight in the interpretation of the review conclusions (2).

5.7 Recommendations on the selection and reporting of model characteristics

The actual aim of this thesis, to provide concrete evidence-based recommendations for the choice of certain experimental parameter specifications, could not be satisfactorily fulfilled based on the results of the meta-reviews, as this would have required the identification of a significant correlation between different model characteristic choices and the experimental results. There, specific choices would have to be beneficial for the use of the model or its reproducibility, e.g. resulting in lower animal mortality, without compromising the modelling quality of SAH. Nevertheless, based on the findings in the reviews of this thesis, it can be strongly advised that selected experimental parameters should be reported in future publications to improve not only the "methods reproducibility" (27, p. 1) but also the value of future reproducibility studies, as more frequent reporting will increase the number of corresponding articles with different specifications of model characteristics and thus the statistical power of the analyses. These recommendations are based on the moderation of model outcomes for cell culture medium glucose levels and SAH perforation filament materials and the identified low reporting frequencies of additional parameters despite their theoretical relevance for the properties of the respective models (2,3,5).

In particular, the reporting of the following model parameters is strongly recommended:

<u>For all models:</u>

- Unambiguous presentation of result estimates and their variances, along with supplementary availability of the underlying raw experimental data, including a clear description of any graphical illustrations

<u>For the U-87 MG in vitro glioblastoma model:</u>

- Source of the U-87 MG cells (company, not scientific colleagues, time of acquisition)
- Implementation and chosen method for cell line authentication
- Cell passage of the cells used in the experiments
- Glucose concentration in the cell culture medium
- Specification of the type of untreated control (medium, saline, drug vehicle, or other)
- Concentration and treatment duration of drugs tested in the model
- Number of independent replications and the underlying technical replications including a

description of the measures taken to achieve independence between the replications

For both investigated SAH mouse models:

- Total number of experimental animals and number of animals per sub-experiment (including information on whether smaller cohorts are subsets of larger cohorts)
- Age, weight, and sex of the experimental animals
- Individual or group housing of the experimental animals
- Choice of anaesthesia method and anaesthetics used during SAH surgeries
- Presence of ICP monitoring during SAH induction surgeries

Specifically for the endovascular perforation SAH model:

- Diameter of the filament used for perforation
- Material of the filament used for perforation
- Tip texture of the filament used for perforation
- Entry site of the filament into the cerebrovascular system
- Location of the endovascular perforation (including methods for location determination)
- Location of vasospasm measurements compared to the location of perforation (ipsi- or contralateral)

Specifically for the cisterna magna injection SAH model:

- Volume of injected blood
- Size of the needle used for injection (diameter)

In addition to these parameters strongly recommended to report in future publications containing the models, the remaining parameters investigated in this work should also be individually considered for reporting due to their potential impact on the models. However, in cases where the space for methodology description is limited and the publication of supplementary material is not possible, the focus should be placed on the listed parameters.

5.8 Perspectives on future research on preclinical research reproducibility

In the past discussion on the reproducibility of preclinic scientific findings, the limited availability of studies on the extent of irreproducibility, especially in preclinical research, was considered a major obstacle to a more precise assessment of the relevance of this problem (23). Thus, the meta-reviews presented in this thesis were conducted to increase the evidence on this issue through retrospective analyses of the reproducibility in selected preclinical models (2,3,5). Although deficits in the reproducibility of basic outcomes of the models were identified, the reasons for irreproducibility could not be satisfactorily established, highlighting the importance of discussing future approaches to reproducibility investigations to subsequently develop effective strategies for improvements where

necessary. In the following sections, such potential approaches are discussed, while all have in common that they would benefit from improved reporting of methodological and outcome parameters in the primary literature.

5.8.1 Elaboration of model-specific moderators of reproducibility

As indicated at various points in this thesis, it is important to identify model-specific variables that affect the reproducibility of model results across studies to sensitise scientists to potential effects of their model set-up on the studied objects. This could be achieved, on the one hand, by meta-reviews, such as those presented in this thesis, and, on the other hand, by experimental work focusing specifically on the influence of these potential variables on model results. Such studies, specifically designed for reproducibility analysis, could systematically vary the suspected reproducibility moderators to measure the effect on the model outcomes.

5.8.2 More efficient use of reproduction attempts

In contrast to meta-reviews, in which many studies are analysed simultaneously to draw conclusions about the across-studies reproducibility, prospective attempts to reproduce individual studies like the RP:CB (238) may also offer valuable insights in reproducibility. However, while one-to-one replication attempts are relatively resource-intensive and do not provide the same comprehensive overview of the currently available literature as meta-reviews, they are expected to be more practice-oriented and could therefore be more helpful in identifying concrete barriers to the reproduction of a particular experiment using a particular preclinical model. To maximise the benefits of often sporadically conducted reproduction-attempting studies, their coordination should be improved. For instance, it has been suggested that a publicly accessible database or journal listing ongoing and finished reproduction studies, accompanied by the corresponding original studies, could be established to increase their recognition (239). This would allow scientists to see directly which models have already proven to be reproducible for studying certain objectives, and which models still require special attention regarding a reproducible performance and presentation of the experiments.

5.8.3 Automation techniques in reproducibility research

With the development of (semi-)automated data analysis techniques, including text mining, machine learning, and algorithm based artificial intelligence, extensive new opportunities for reproducibility analysis are expected (240). This could enable the processing of large volumes of published research data that could not be handled by purely human analysts (240). For example, there are already promising approaches to supported automated literature screening in systematic reviews (241). In the future, it may be possible to conduct routine retrospective meta-analyses using automated data extraction and analysis techniques, as well as prospective reproduction studies using experimental robots, especially in the in vitro field, to assess the reproducibility of findings (240). To further support automated analysis, it may be necessary to report key model parameters in tabular form rather than narrative text, which would also enhance the visibility of methodological elements and facilitate verification of compliance with reporting guidelines. Regarding experimental robots, the automated

execution of in vitro experiments could lead to more accurate performance and may reduce bias due to the frequent absence of randomisation and blinding (242). In addition, their implementation may also reduce the cost of reproduction experiments (240). Overall, however, the true potential of these techniques and the remaining degree of human assistance required is currently difficult to estimate, as the testing phase of automation has so far been limited to less complex experimental settings, such as the analysis of the influence of added drugs on the gene expression of cells (240). Nonetheless, the current trend towards automation and the use of ever-increasing amounts of research data should not be neglected in reproducibility studies, especially as important scientific findings could be obtained as a by-product of reproducibility investigations (240), since they are essentially attempts to identify systematic patterns and reasons for variation in the studied field of science.

5.8.4 Randomisation, blinding, and sample size calculation in preclinical research

Randomisation, blinding, and sample size calculations are well established in clinical trials to reduce the risk of bias and to estimate necessary cohort sizes, and it can be assumed that these features are also relevant for in vitro and in vivo studies, especially in comparative intervention experiments, which have a similar basic structure to clinical trials. However, both the reviews in this thesis as well as external investigations indicated that these methods are very rarely used in the preclinic, especially in the in vitro setting (2,3,5,243,244), despite recommendations for their implementation (8,158,245–247). Due to the nature of the investigated models, the analysis of the impact of these features was limited, for example, as the SAH induction surgery is difficult to perform without the surgeon's knowledge of group allocation. Nevertheless, attention should be paid to blinded and randomised incubation of cells and housing of animals to minimise confounding factors and sources of irreproducibility, because even if cells and animals are derived from a presumably more homogeneous population than humans due to more controllable environmental conditions, (un)conscious differences in their treatment by the researchers may have systematically influenced their characteristics in response to the tested interventions (245). For the measurement of in vitro data, automated procedures should be preferred where available to minimise the risk of bias due to a lack of blinding. Regarding sample size calculations under in vitro conditions, authors often reported the performance of three independent replications as an apparent standard, but especially because of the difficulty in defining the exact experimental unit, the methods chosen to achieve independence should be sufficiently explained, as mentioned earlier (5). For in vivo studies, a prior calculation of needed experimental animal group sizes is useful to limit the number of animals to the minimum in accordance with the “principles of the 3Rs [sic]” (22, p. 1) and should be reported in respective publications (166).

However, before providing concrete recommendations on the implementation of randomisation, blinding, and sample size calculation in preclinical research, their impact on reproducibility and research quality should be investigated more closely to justify the additional effort associated with their implementation. Nevertheless, relatively recent observations of in vivo stroke models have shown that the implementation of randomisation and blinding is increasing, with a current prevalence of approximately 40 % in this specific area (248).

5.9 Perspectives on the enhancement of reporting quality

As mentioned above, improvements in the reporting quality of preclinical research are urgently needed to strengthen "methods reproducibility" (27, p. 1), facilitate the interpretation of results, and improve the conditions for future reproducibility analyses. Approaches that may be used to achieve these improvements are discussed below. It should be noted that although more detailed reporting of methodology and results initially requires additional effort, this may ultimately be offset by more efficient scientific processes. Furthermore, the existence of articles with a relatively good reporting quality (2,3,5) demonstrated that accurate and comprehensive reporting in peer-reviewed research articles is possible.

5.9.1 Reporting guideline adherence as a requirement for publication

One apparent approach is the introduction of mandatory reporting guidelines, which authors would have to follow to be eligible for publication. However, as mentioned in the introduction of this thesis, the dissemination of reporting guidelines in the preclinical field, especially for in vitro research, is rather sparse, so that first guidelines for different areas of preclinical research would have to be increasingly developed by expert groups. In addition, existing reporting guidelines need to be enforced more strictly, as the example of the ARRIVE guideline shows (40). In the applicable field of animal studies, the SAH model reviews (2,3) as well as external data (50,249,250) showed that the reporting of the required items has not yet improved to the desired extent intended by the guideline creators.

During publication processes of peer-reviewed research articles, reporting guidelines could be enforced in the form of checklists that need to be completed by both authors and reviewers. In these, authors would need to declare where in the article the relevant parameters are reported, which would then be checked for correctness by peer reviewers. To ensure that the assessment is carried out with the necessary accuracy, comments on compliance for each required parameter could be published along with the articles, so that both parties would be motivated to ensure proper compliance. It is also imaginable that compliance with reporting guidelines in past studies could become a requirement for approval to conduct future studies, like required ethical votes. This would be a more rigorous approach, requiring universities and other research institutions to regularly check whether studies meet defined reporting standards. If this is repeatedly not the case, a follow-up seminar could be required to raise awareness of the importance of adequate reporting before further studies are approved. However, a major challenge would be the simultaneous global implementation of reporting guideline adherence checks to maintain equal opportunities for researchers. Reviewing the quality of reporting would also require additional resources from the institutions, and it would be necessary to ensure that the review boards meet quality criteria to ensure comparable standards. In addition to these two approaches, which would introduce additional bureaucracy and hurdles and thus may be unpopular in large parts of the research community, another strategy might be the implementation of an intuitive indicator of reporting quality. Scientific articles could be given a label or score indicating the extent to which they meet the required reporting standards. Unlike the previous two proposals, this would not deny publication to any

research group, but by making reporting quality easily visible, it would create an incentive to report research properly to maintain the reputation of one's own research.

For all approaches, the existence of appropriate reporting guidelines for specific scientific fields is a prerequisite before their implementation can be widely enforced. Furthermore, it needs to be determined who assesses the reporting quality of studies. If this is done by the peer reviewers of the journals, a potential conflict of interest could arise, as the reviewers may have an interest in having the articles in their journals attested to be of a higher reporting quality on average, to raise the perceived quality of the journal. Therefore, external reporting assessments by blinded reviewers could be useful but would require additional effort.

5.9.2 Elimination of word limits in research articles

Another approach to promote reporting of preclinical research may be the removal of word limits for the methods section of scientific publications to avoid discouraging scientific authors from presenting a full description of their methodology. However, as the capacity of scientific journals is of course limited, alternatives should be offered for exceeding method descriptions, such as additional, more detailed descriptions in freely accessible online supplements. Positive examples of dealing with the length of the method description include the journal *Nature*, where authors are actively encouraged to provide an extended version of methods outside the main research article on their own protocol platform (251), as well as *PLOS*, which completely dispenses with a hard word limits (252). At the same time, it is important that descriptions of the methodology remain concise, providing the important aspects for reproducing the experimental set-up and understanding the presented results.

5.9.3 Raising awareness of the importance of accurate reporting

To improve reporting in a sustainable way, it will be important to raise the general awareness among researchers of the importance of adequate reporting for the quality and value of their research and as a prerequisite for reproducibility and thus scientific progress. Concrete actions that could be taken to raise awareness may include the following approaches:

1. Organisation of workshops in which, for example, a fictitious study is presented in different versions of increasing reporting qualities, while the participants are asked to judge the reproducibility and informative value of the presented data which may lead to an eye-opening effect with self-realisation of the value of adequate reporting.
2. Similar to the first approach, positive personal experiences of researchers reading articles by other authors with reproducible and transparent reporting may also motivate researchers to improve their reporting, because of a better understanding of the added value of sufficient reporting.
3. Reporting guidelines should include a detailed rationale for the inclusion of the required parameters and their potential relevance to the experimental results, so that researchers can understand the benefits of reporting and thus may be more likely to adhere to them.
4. During the journal peer review process, constructive feedback with concrete suggestions from reviewers on how to improve reporting may also help to elevate reporting standards.

5.9.4 Requirement of pre-study protocols

Another strategy to improve the quality of reporting of preclinical studies may be an increased demand for the publication of study protocols prior to the execution of experiments similar to clinical research, where a priori study protocols are common (30). With this approach, the reporting of important study parameters could be outsourced, as the protocols should contain a detailed description of the planned experiments, including the used models with their specifications, sample sizes, experimental animals, and statistical analyses. A reference to the protocol with details and explanations of any deviations in the conduct of the experiments should be included in the final publication (17,253). As a side effect, the requirement to publish pre-study protocols may encourage authors to reflect more on a reproducible choice of model parameters. However, it is possible that deviations from protocols may not always be fully reported, so that incorrect model parameters could be assumed for the interpretation of study results.

5.10 Perspectives on the enhancement of reproducibility in preclinical research

In addition to strategies to improve the reporting of preclinical research, proactive approaches to promote reproducibility should also be pursued, as the need for such measures has been demonstrated, not least in the reviews presented in this thesis (2,5). In the following, selected concepts for such improvements are discussed.

5.10.1 Raising awareness of the importance of reproducibility

While the general awareness of irreproducibility as a problem in science seems to be present, with, for example, around 90 % of respondents in a survey of 1,576 scientists stating that there are problems with reproducibility in science, the deeper understanding of the consequences of irreproducibility and its drivers may be less widespread, as only 41 % of respondents in the same survey have taken concrete actions to improve the reproducibility of their research (31). In contrast, in clinical research, the awareness regarding reproducibility, rigor, and transparency in studies has increased, not least due to numerous initiatives over the last decades (254,255). In the preclinical setting, however, this process has slowly commenced but is still far from the standards in the clinic (254,255). Therefore, as for the reporting quality, the awareness of the importance of reproducibility among scientists should also be strengthened to achieve sustainable improvements. The approaches for raising awareness are similar to those for improving reporting, i.e. through specific training, clear explanations of the importance and benefits of reproducible research, constructive personal feedback on how to improve reproducibility, and personal experience of the benefits of reproducibility. For instance, an improved interpretation of published studies due to a stronger focus on reproducible conduct and reporting of the experiments could avoid redundant work and provide valuable input to one's own projects, so the initial extra effort of more precise documentation and disclosure of raw data and analyses, as well as extra effort through reproducibility tests, may be outweighed by the long-term benefits of more reproducible findings.

5.10.2 Database of preclinical models

Due to the large variety of preclinical models with multiple options in different configurations per disease, reproducibility could also be increased by the establishment of a model database, as it has already been suggested for in vivo SAH models (201). In such a database of in vitro and in vivo models, categorised according to the modelled diseases, research articles could be tagged with the used models, providing a direct overview of relevant publications per model. In addition, for each model, selected model specifications and results for key endpoints such as cell growth inhibition and mortality could be displayed, together with overall statistics on the frequency of model parameter choices, to provide a general overview of the model heterogeneity. Such a database would facilitate the selection of an appropriate model and the optimal configuration of its specifications for specific research questions. Furthermore, future reproducibility analysis would profit from a concise presentation of primary results along with model specifications. However, the establishment and maintenance of such a database would require considerable resources and funding. Ideally, journals could require authors to provide basic data on the model, its specifications, and primary results during the editorial process, so that these could be relatively easily incorporated into the database at the time of publication. A further challenge would be the selection of parameters displayed for each model given the diversity of existing models. The selection would initially need to be made by experts in the field but should later be open to external suggestions. Moreover, as the displayed results would represent only a small fraction of each study, it must be emphasised that the full publication should always be considered to avoid prejudging and misleading conclusions based on the selective presentation of studies in the database.

5.10.3 Obligatory reproduction attempt before publication

As earlier mentioned, reproducibility may also be improved by the introduction of mandatory reproducibility testing before the publication of studies. In these, the experiments are reproduced based on the methods described in the articles to assess whether similar results to the published data can be observed. As this would indicate the “methods reproducibility” (27, p. 1) and “results reproducibility” (27, p. 1), the confidence in the validity of studies may be increased. Moreover, they could also serve as a practical quality control, for example by highlighting inappropriate statistical analyses or an excessive influence of bias. In addition, authors of original studies might provide a more complete description of their methods and results and elevate internal reproducibility checks to avoid negative reproducibility claims about their research. The reproducibility tests could be done either by the authors of the study themselves or by external researchers. In this way, the current *peer-reviewed* standard for research articles could become a new *peer-reviewed and reproduced* standard. They also could be made optional rather than mandatory, so that reproduction attempts would only have to be carried out if the authors wished to add such a *reproduced* label to their article. The reproducibility check should be reported in a condensed form alongside the original study, with a full description of the reproduced methods and results, and a statement about the challenges and successes of the reproduction.

If external researchers are required to perform the reproduction attempts, which would tend to be more meaningful than if the original researchers reproduce the experiments they are already familiar with as they are therefore not necessarily dependent on the information provided in the publication, they

will probably need to be recruited by the publishing journals. For this purpose, researchers interested in performing reproduction experiments should contact the journal with a declaration of their respective expertise, so that the journal can assign them suitable studies. However, as this would involve considerable cost and effort, incentives would have to be created for reproducing scientists. This could be achieved by explicitly acknowledging the researchers who performed the reproduction attempts, for example by including them in the reference, as in the following fictional example: *Max Mustermann et al, reproduced by Tommy Atkins et al.* In addition, financial compensation for the costs associated with reproduction experiments would probably need to be offered. Controversially, both too high and too low motivation to perform reproduction attempts may cause problems in their implementation. On the one hand, scientists who decide to undertake a reproduction attempt may be more motivated to publish negative reproducibility findings, as this would presumably attract more attention and put themselves in a seemingly superior position as correctors of the original findings (233). On the other hand, the motivation may be rather low as new and innovative research tends to receive more attention and is associated with higher financial and reputational benefits (30). In any case, it would be necessary to establish a review process for the reproduction experiments themselves, to ensure that they are carried out according to scientific standards, and in particular to avoid prematurely degrading the potential value of the checked original research. In general, if it turns out that the findings of a study are completely or partially irreproducible, the study should nevertheless be published together with the results of the reproduction attempt, ideally with a joint statement by the original and reproduction authors on possible reasons for the diverging findings and the resulting implications for further research on the specific topic.

A disadvantage of reproduction attempts would be the considerable additional effort involved and a further prolongation of the already lengthy publication process which may prevent important new findings from being published in a timely manner. To address this, preliminary versions of studies could be published with a note that a reproducibility check is pending. Furthermore, despite the incentives mentioned above, it may still be difficult to recruit a sufficient number of researchers for the reproducibility checks. And even if they were found, they would not be available for their own innovative research during the reproduction experiments, which could slow down scientific progress. However, a structured assessment of the reproducibility of published scientific data could in turn lead to an overall more efficient scientific progress and compensate for the additional efforts, as future research based on robust published data is potentially more successful if these are tested for their reproducibility beforehand. In addition, the reproducibility checks would be better coordinated rather than carried out in parallel and only partially by different research groups, which could further improve efficiency. Another potentially limiting aspect that needs to be considered is that scientists may be discouraged from publishing innovative research, as this is more likely to be less reproducible than research using established methods.

Overall, the introduction of mandatory or optional reproducibility checks could be a useful element to enhance the reproducibility in preclinical science. However, it is currently difficult to assess

the acceptance by the scientific community regarding the involved increased effort. As a compromise and first step, more attention could be given to the implementation of reproducibility assessments by the publishing researchers themselves and in cases of doubtful results, external reproducibility tests may be performed selectively. Besides reproducibility, it is also essential to examine the generalisability of the results, which primarily involves testing the investigated hypotheses on different models at different levels of research, i.e. in vitro, in vivo, and in humans (30).

5.10.4 Online platform to promote interaction between researchers

To enable as many researchers as possible to learn from the difficulties encountered in reproduction attempts of published data, effective solution strategies should be made publicly available. For this purpose, an online platform may be developed to allow researchers to contact authors of original publications when they encounter difficulties in reproductions. There, the authors would discuss publicly with the questioners to identify possible reasons for an unsuccessful reproduction attempt, e.g. deviations in the experimental setting or measurement methods. Moreover, the platform could be used to coordinate collaborative reproduction efforts by multiple research groups to efficiently allocate resources. While platforms such as the *Open Science Framework*, *GitHub,* and *Zenodo* (256–258) exist that allow the publication of study protocols, raw data, and statistical analysis to increase transparency in science, the difference to the platform proposed here is the focus on solving and benefiting from difficulties in reproduction attempts through public interaction between researchers. Potentially, such a communication feature could be integrated into one of the existing platforms, which might reduce the cost of implementing this approach to an increased reproducibility. However, it is important to stress that this feature is not intended to replace scientific journals with their own peer-review processes. Instead, articles published in journals could have a direct link to the listing of the article on the proposed platform. Given the generally low response rates when original authors are contacted for additional information on the applied methodology (5,259,260), a point system for quick and informative responses could additionally encourage constructive feedback. Furthermore, moderation of the public discussion area would be necessary to maintain scientific standards, which would again require additional effort and funded resources.

5.10.5 Reproducibility improvements do not aim for total standardisation

It is important to emphasise that the current and proposed initiatives to improve reporting and reproducibility should not aim to achieve a complete standardisation of models, where every research group uses the exact same experimental set-up and all variance in a preclinical model is lost. Such a situation would carry the risk that the obtained results might only be valid under those specific conditions. Thus, while the internal reproducibility within the model would be increased by a standardisation, the external reproducibility, generalisability, and transferability to the clinic would likely be severely reduced (195,261). What should be standardised, however, is the way in which preclinical research is reported so that the impact of different experimental parameters can be assessed. Subsequently, if the analyses show that a certain choice of a model characteristic leads to a significantly higher reproducibility or a lower frequency of adverse outcomes, e.g., higher mortality, without affecting

the modelling quality of the disease, this specific parameter may be standardised. Overall, however, the heterogeneity of research, especially in the preclinical stage, should not be diminished, as it is often through this heterogeneity that valuable observations can be made, e.g. regarding treatment response in different population subgroups. To facilitate such discoveries, systematic heterogenisation is an approach in which experimental parameters are intentionally varied to assess their impact on study outcomes (195). This way, more reproducible science does not intend that all experiments should be carried out in exactly the same way, but that variations should be accurately documented so that the resulting variation in results, i.e. irreproducibility, can be understood.

5.11 Conclusion

In conclusion, the objectives of this thesis were partially met. A meta-analytic quantification of irreproducibility was achieved for both the U-87 MG model and the endovascular perforation SAH model, for which significant variance between studies was observed for the TMZ-induced inhibition of cell viability and the mouse mortality after SAH, respectively (2,5). In contrast, no significant heterogeneity was found for the SAH severity grade in the endovascular perforation SAH model (2). However, no meta-analysis of the reproducibility of basic model results could be performed for the cisterna magna injection SAH model and the outcome of large artery vasospasm in the endovascular perforation SAH model due to the limited number of underlying data in the primary literature, although the available literature also suggested meaningful variance (2,3). Despite the fact that the model parameters glucose concentration in the U-87 MG cell culture medium and the material of the filament used for endovascular perforation were shown to affect the model outcomes and thus their reproducibility, none of the different specifications of these parameters was identified as offering significantly improved conditions for reproducibility or advantages for the modelling quality of the respective disease (2,5). Therefore, it was not reasonable to provide specific recommendations for the choice of model characteristics. The main suspected reason for the limited statistical power of the meta-reviews and the associated difficulties in identifying sources of irreproducibility was the low frequency of reporting of several key model parameters in the underlying included original research articles (2,5). This severely limited the methodological reproducibility of the studies and often prevented a clear interpretation of the presented results. Moreover, this highlighted the need to first focus on improving the reporting of preclinical neurosurgical research in order to improve the conditions for future reproducibility analyses in this field. The strategies discussed to enhance reporting and reproducibility are diverse and may include reporting practices, raising awareness among researchers, pre-study protocols, reproduction attempts, model databases, as well as interaction platforms to address specific reproducibility challenges.

Overall, the findings on the limited reporting quality and reproducibility of selected preclinical neurosurgical models underline the need for action to improve these research quality attributes, and at the same time can serve as an evidence-based foundation for the development of further initiatives to improve towards a more transparent and reproducible preclinical research, with the ultimate aim of increasing the rate of successful translations of preclinical findings into human clinical application and

an improved prognosis for these two exemplary very severe diseases. For this aim, the approaches discussed in this thesis are likely to be applicable beyond the investigated models and the field of preclinical neurosurgical research.

6 References

1. Alpdogan S, Sander T, Zhang R, Khan D, Li X, Zhou H, et al. Meta review on perforation model of subarachnoid haemorrhage in mice – filament material as a moderator of mortality. In German Medical Science GMS Publishing House; 2022. p. DocP084.

2. Alpdogan S, Sander T, Zhang R, Khan D, Li X, Zhou H, et al. Meta-review on Perforation Model of Subarachnoid Hemorrhage in Mice: Filament Material as a Possible Moderator of Mortality. Translational Stroke Research. 2022 Nov 23;

3. Alpdogan S, Li K, Sander T, Cornelius JF, Muhammad S. Cisterna Magna Injection Mouse Model of Subarachnoid Hemorrhage (SAH): A Systematic Literature Review of Preclinical SAH Research. J Exp Neurol. 2023 Mar 24;Volume 4(Issue 1):11–20.

4. Sander T, Ghanawi J, Wilson E, Muhammad S, MacLeod M, Kahlert U. Reproducibility of in vitro glioma research exemplified by the U-87 MG model. In German Medical Science GMS Publishing House; 2022. p. DocP157.

5. Sander T, Ghanawi J, Wilson E, Muhammad S, Macleod M, Kahlert UD. Meta-analysis on reporting practices as a source of heterogeneity in in vitro cancer research. BMJ Open Sci. 2022 Jun 1;6(1):e100272.

6. Fanelli D. Is science really facing a reproducibility crisis, and do we need it to? Proceedings of the National Academy of Sciences. 2018 Mar 13;115(11):2628–31.

7. The Economist. How science goes wrong. The Economist [Internet]. 2013 Oct 19 [cited 2024 Jun 1]; Available from: https://www.economist.com/leaders/2013/10/21/how-science-goes-wrong

8. Begley CG, Ellis LM. Raise standards for preclinical cancer research. Nature. 2012 Mar;483(7391):531–3.

9. Jarvis MF, Williams M. Irreproducibility in Preclinical Biomedical Research: Perceptions, Uncertainties, and Knowledge Gaps. Trends in Pharmacological Sciences. 2016 Apr 1;37(4):290–302.

10. Pashler H, Harris CR. Is the Replicability Crisis Overblown? Three Arguments Examined. Perspect Psychol Sci. 2012 Nov 1;7(6):531–6.

11. Collins FS, Tabak LA. NIH plans to enhance reproducibility. Nature. 2014 Jan 30;505(7485):612–3.

12. National Institutes of Health (NIH). Enhancing Reproducibility through Rigor and Transparency [Internet]. 2023 [cited 2024 Jun 1]. Available from: https://grants.nih.gov/policy/reproducibility/index.htm

13. ASCB Data Reproducibility Task Force. Difficulty Reproducing Research Data Examined by ASCB Task Force [Internet]. ASCB. 2014 [cited 2024 Jun 1]. Available from: https://www.ascb.org/science-policy-public-outreach/advocacy-policy/ascb-examines-difficulty-reproducing-research-data/

14. Begley C. Glenn, Ioannidis John P.A. Reproducibility in Science. Circulation Research. 2015 Jan 2;116(1):116–26.

15. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. PLOS Biology. 2015 Jun 9;13(6):e1002165.

16. Global Biological Standards Institute (GBSI). The Case for Standards in Life Science Research Seizing Opportunities at a Time of Critical Need. 2013 [cited 2021 Apr 9]; Available from: https://www.academia.edu/35147125/The_Case_for_Standards_in_Life_Science_Research_Seizing_Opportunities_at_a_Time_of_Critical_Need

17. Hirsch C, Schildknecht S. In Vitro Research Reproducibility: Keeping Up High Standards. Front Pharmacol. 2019;10.

18. 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 Sep;10(9):712–712.

19. Wen H, Wang HY, He X, Wu CI. On the low reproducibility of cancer studies. Natl Sci Rev. 2018 Sep;5(5):619–24.

20. Errington TM, Mathur M, Soderberg CK, Denis A, Perfito N, Iorns E, et al. Investigating the replicability of preclinical cancer biology. Pasqualini R, Franco E, editors. eLife. 2021 Dec 7;10:e71601.

21. Merton RK. The Sociology of Science. The University of Chicago Press; 1973.

22. NC3Rs. The 3Rs [Internet]. [cited 2024 Jun 2]. Available from: https://www.nc3rs.org.uk/who-we-are/3rs

23. Kraus WL. Editorial: Do You See What I See?: Quality, Reliability, and Reproducibility in Biomedical Research. Mol Endocrinol. 2014 Mar;28(3):277–80.

24. Tonzani S, Fiorani S. The STAR Methods way towards reproducibility and open science. iScience. 2021 Apr 1;24(4):102137.

25. Lazic SE. Higher Education from Cambridge University Press. Cambridge University Press; 2016 [cited 2023 Apr 21]. Experimental Design for Laboratory Biologists: Maximising Information and Improving Reproducibility. Available from: https://www.cambridge.org/highereducation/books/experimental-design-for-laboratory-biologists/31C1A347D0ADB25226D7220A99C0EF56

26. Plesser HE. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. Frontiers in Neuroinformatics. 2018;11.

27. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? Science Translational Medicine. 2016 Jun 1;8(341):341ps12-341ps12.

28. Stark PB. Before reproducibility must come preproducibility. Nature. 2018 May 24;557(7707):613–613.

29. Baker M. Biotech giant publishes failures to confirm high-profile science. Nature. 2016 Feb 1;530(7589):141–141.

30. Fazilaty H. Assessing reproducibility of the core findings in cancer research. iScience. 2022 Sep 13;25(10):105125.

31. Baker M. 1,500 scientists lift the lid on reproducibility. Nature News. 2016 May 26;533(7604):452.

32. Mobley A, Linder SK, Braeuer R, Ellis LM, Zwelling L. A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic. Arakawa H, editor. PLoS ONE. 2013 May 15;8(5):e63221.

33. Williams M. Reagent Validation to Facilitate Experimental Reproducibility. Current Protocols in Pharmacology. 2018;81(1):e40.

34. An G. The Crisis of Reproducibility, the Denominator Problem and the Scientific Role of Multi-scale Modeling. Bull Math Biol. 2018 Dec 1;80(12):3071–80.

35. 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 Aug 28;3(8):e3081.

36. Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: Progress and priorities. F1000Res. 2017 May 2;6:604.

37. Hartung T, Vries R de, Hoffmann S, Hogberg HT, Smirnova L, Tsaioun K, et al. Toward Good In Vitro Reporting Standards. ALTEX. 2019 Jan 8;36(1):3–17.

38. Centre for Statistics in Medicine (CSM), NDORMS, University of Oxford. EQUATOR Network | Enhancing the QUAlity and Transparency Of Health Research [Internet]. 2023 [cited 2024 Jun 1]. Available from: https://www.equator-network.org/

39. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLOS Biology. 2010 Jun 29;8(6):e1000412.

40. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. Boutron I, editor. PLoS Biol. 2020 Jul 14;18(7):e3000410.

41. Saver JL, Albers GW, Dunn B, Johnston KC, Fisher M. Stroke Therapy Academic Industry Roundtable (STAIR) Recommendations for Extended Window Acute Stroke Therapy Trials. Stroke. 2009 Jul 1;

42. Savitz SI, Chopp M, Deans R, Carmichael ST, Phinney D, Wechsler L. Stem Cell Therapy as an Emerging Paradigm for Stroke (STEPS) II. Stroke. 2011 Mar;42(3):825–9.

43. Lapchak PA, Zhang JH, Noble-Haeusslein LJ. RIGOR Guidelines: Escalating STAIR and STEPS for Effective Translational Research. Transl Stroke Res. 2013 Jun 1;4(3):279–85.

44. Goursaud S, de Lizarrondo S, Grolleau F, Chagnot A, Agin V, Maubert E, et al. Delayed Cerebral Ischemia After Subarachnoid Hemorrhage: Is There a Relevant Experimental Model? A Systematic Review of Preclinical Literature. FRONTIERS IN CARDIOVASCULAR MEDICINE. 2021 Nov 15;8.

45. Grüter BE, Croci D, Schöpf S, Nevzati E, d'Allonzo D, Lattmann J, et al. Systematic Review and Meta-analysis of Methodological Quality in In Vivo Animal Studies of Subarachnoid Hemorrhage. Transl Stroke Res. 2020 Dec;11(6):1175–84.

46. Macleod, Collings AM, Graf C, Kiermer V, Mellor D, Swaminathan S, et al. The MDAR (Materials Design Analysis Reporting) Framework for transparent reporting in the life sciences. PNAS [Internet]. 2021 Apr 27 [cited 2024 Jun 1];118(17). Available from: https://www.pnas.org/content/118/17/e2103238118

47. Organisation for Economic Co-operation and Development (OECD). Guidance Document on Good In Vitro Method Practices (GIVIMP) | en | OECD [Internet]. 2018 [cited 2024 Jun 1]. Available from: https://www.oecd.org/env/guidance-document-on-good-in-vitro-method-practices-givimp-9789264304796-en.htm

48. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. The Lancet. 2014 Jan 18;383(9913):267–76.

49. Hackam DG, Redelmeier DA. Translation of Research Evidence From Animals to Humans. JAMA. 2006 Oct 11;296(14):1727–32.

50. Macleod MR, McLean AL, Kyriakopoulou A, Serghiou S, Wilde A de, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLOS Biology. 2015 Oct 13;13(10):e1002273.

51. 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 Oct 15;139(2):243.

52. Worp HB van der, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can Animal Models of Disease Reliably Inform Human Studies? PLOS Medicine. 2010 Mar 30;7(3):e1000245.

53. Else H. Radical open-access plan could spell end to journal subscriptions. Nature. 2018 Sep 4;561(7721):17–8.

54. Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg LS, et al. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biology. 2016 May 12;14(5):e1002456.

55. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proceedings of the National Academy of Sciences. 2018 Mar 13;115(11):2600–6.

56. Kahlert UD, Koch K, Suwala AK, Hartmann R, Cheng M, Maciaczyk D, et al. The effect of neurosphere culture conditions on the cellular metabolism of glioma cells. Folia Neuropathol. 2015;53(3):219–25.

57. van Lieshout JH, Marbacher S, Muhammad S, Boogaarts HD, Bartels RHMA, Dibué M, et al. Proposed Definition of Experimental Secondary Ischemia for Mouse Subarachnoid Hemorrhage. Transl Stroke Res. 2020 Oct 1;11(5):1165–70.

58. Brown NF, Ottaviani D, Tazare J, Gregson J, Kitchen N, Brandner S, et al. Survival Outcomes and Prognostic Factors in Glioblastoma. Cancers (Basel). 2022 Jun 28;14(13):3161.

59. Suzuki H, Nakano F. To Improve Translational Research in Subarachnoid Hemorrhage. Transl Stroke Res. 2018 Feb 1;9(1):1–3.

60. Robertson FL, Marqués-Torrejón MA, Morrison GM, Pollard SM. Experimental models and tools to tackle glioblastoma. Disease Models & Mechanisms. 2019 Sep 6;12(dmm040386).

61. Pontén J, Macintyre EH. Long Term Culture of Normal and Neoplastic Human Glia. Acta Pathologica Microbiologica Scandinavica. 1968;74(4):465–86.

62. Allen M, Bjerke M, Edlund H, Nelander S, Westermark B. Origin of the U87MG glioma cell line: Good news and bad news. Science Translational Medicine. 2016 Aug 31;8(354):354re3-354re3.

63. HTB-14™, ATCC®. U-87 MG ATCC ® HTB-14™ Homo sapiens brain Likely glioblastom [Internet]. 2021 [cited 2024 Jun 1]. Available from: https://www.lgcstandards-atcc.org/Products/All/HTB-14.aspx?geo_country=de

64. Lee CH, Jung KW, Yoo H, Park S, Lee SH. Epidemiology of Primary Brain and Central Nervous System Tumors in Korea. J Korean Neurosurg Soc. 2010 Aug;48(2):145–52.

65. Pouchieu C, Gruber A, Berteaud E, Ménégon P, Monteil P, Huchet A, et al. Increasing incidence of central nervous system (CNS) tumors (2000–2012): findings from a population based registry in Gironde (France). BMC Cancer. 2018 Dec;18(1):1–13.

66. Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014–2018. Neuro Oncol. 2021 Oct 5;23(Suppl 3):iii1–105.

67. Burnet NG, Jefferies SJ, Benson RJ, Hunt DP, Treasure FP. Years of life lost (YLL) from cancer is an important measure of population burden – and should be considered when allocating research funds. Br J Cancer. 2005 Jan 31;92(2):241–5.

68. McKinnon C, Nandhabalan M, Murray SA, Plaha P. Glioblastoma: clinical presentation, diagnosis, and management. BMJ. 2021 Jul 14;374:n1560.

69. Larjavaara S, Mäntylä R, Salminen T, Haapasalo H, Raitanen J, Jääskeläinen J, et al. Incidence of gliomas by anatomic location. Neuro Oncol. 2007 Jul;9(3):319–25.

70. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. 2021 Jun 29;23(8):1231–51.

71. Tan AC, Ashley DM, López GY, Malinzak M, Friedman HS, Khasraw M. Management of glioblastoma: State of the art and future directions. CA: A Cancer Journal for Clinicians. 2020;70(4):299–312.

72. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma. New England Journal of Medicine. 2005 Mar 10;352(10):987–96.

73. Herrlinger U, Tzaridis T, Mack F, Steinbach JP, Schlegel U, Sabel M, et al. Lomustine-temozolomide combination therapy versus standard temozolomide therapy in patients with newly diagnosed glioblastoma with methylated MGMT promoter (CeTeG/NOA–09): a randomised, open-label, phase 3 trial. The Lancet. 2019 Feb 16;393(10172):678–88.

74. Krigers A, Pinggera D, Demetz M, Kornberger LM, Kerschbaumer J, Thomé C, et al. The Routine Application of Tumor-Treating Fields in the Treatment of Glioblastoma WHO° IV. Front Neurol. 2022 Jun 16;13:900377.

75. Maschio M, Dinapoli L, Sperati F, Pace A, Fabi A, Vidiri A, et al. Levetiracetam monotherapy in patients with brain tumor-related epilepsy: seizure control, safety, and quality of life. J Neurooncol. 2011 Aug 1;104(1):205–14.

76. Krex D, Klink B, Hartmann C, von Deimling A, Pietsch T, Simon M, et al. Long-term survival with glioblastoma multiforme. Brain. 2007 Oct;130(Pt 10):2596–606.

77. Rodríguez-Camacho A, Flores-Vázquez JG, Moscardini-Martelli J, Torres-Ríos JA, Olmos-Guzmán A, Ortiz-Arce CS, et al. Glioblastoma Treatment: State-of-the-Art and Future Perspectives. Int J Mol Sci. 2022 Jun 29;23(13):7207.

78. Torsvik A, Stieber D, Enger PØ, Golebiewska A, Molven A, Svendsen A, et al. U-251 revisited: genetic drift and phenotypic consequences of long-term cultures of glioblastoma cells. Cancer Medicine. 2014;3(4):812–24.

79. Capes-Davis A, Bairoch A, Barrett T, Burnett EC, Dirks WG, Hall EM, et al. Cell Lines as Biological Models: Practical Steps for More Reliable Research. Chem Res Toxicol. 2019 Sep 16;32(9):1733–6.

80. Coecke S, Balls M, Bowe G, Davis J, Gstraunthaler G, Hartung T, et al. Guidance on Good Cell Culture Practice: A Report of the Second ECVAM Task Force on Good Cell Culture Practice. Altern Lab Anim. 2005 Jun 1;33(3):261–87.

81. International Cell Line Authentication Committee. ANSI/ATCC ASN-0002-2011 - Authentication of Human Cell Lines: Standardization of STR Profiling [Internet]. 2012 [cited 2024 Jun 1]. Available from: https://webstore.ansi.org/standards/atcc/ansiatccasn00022011

82. Weeber F, Ooft SN, Dijkstra KK, Voest EE. Tumor Organoids as a Pre-clinical Cancer Model for Drug Discovery. Cell Chemical Biology. 2017 Sep 21;24(9):1092–100.

83. Hafner M, Niepel M, Chung M, Sorger PK. Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. Nat Methods. 2016 Jun;13(6):521–7.

84. Niepel M, Hafner M, Mills CE, Subramanian K, Williams EH, Chung M, et al. A Multi-center Study on the Reproducibility of Drug-Response Assays in Mammalian Cell Lines. cels. 2019 Jul 24;9(1):35-48.e5.

85. McKee TJ, Komarova SV. Is it time to reinvent basic cell culture medium? American Journal of Physiology-Cell Physiology. 2017 May;312(5):C624–6.

86. Samanthi. Compare the Difference Between Similar Terms. 2018 [cited 2023 Mar 16]. Difference Between RPMI and DMEM. Available from: https://www.differencebetween.com/difference-between-rpmi-and-vs-dmem/

87. Poon MTC, Bruce M, Simpson JE, Hannan CJ, Brennan PM. Temozolomide sensitivity of malignant glioma cell lines – a systematic review assessing consistencies between in vitro studies. BMC Cancer. 2021 Nov 18;21(1):1240.

88. Bao Z, Chen K, Krepel S, Tang P, Gong W, Zhang M, et al. High Glucose Promotes Human Glioblastoma Cell Growth by Increasing the Expression and Function of Chemoattractant and Growth Factor Receptors. Translational Oncology. 2019 Sep 1;12(9):1155–63.

89. Bielecka-Wajdman AM, Ludyga T, Smyk D, Smyk W, Mularska M, Świderek P, et al. Glucose Influences the Response of Glioblastoma Cells to Temozolomide and Dexamethasone. Cancer Control. 2022 Feb 26;29:10732748221075468.

90. Relier S, Yazdani L, Ayad O, Choquet A, Bourgaux JF, Prudhomme M, et al. Antibiotics inhibit sphere-forming ability in suspension culture. Cancer Cell Int. 2016 Feb 12;16:6.

91. Pilgrim CR, McCahill KA, Rops JG, Dufour JM, Russell KA, Koch TG. A Review of Fetal Bovine Serum in the Culture of Mesenchymal Stromal Cells and Potential Alternatives for Veterinary Medicine. Frontiers in Veterinary Science. 2022;9.

92. Hatzis C, Bedard PL, Birkbak NJ, Beck AH, Aerts HJWL, Stern DF, et al. Enhancing Reproducibility in Cancer Drug Screening: How Do We Move Forward? Cancer Res. 2014 Aug 1;74(15):4016–23.

93. Zhang L, Yu H, Yuan Y, Yu JS, Lou Z, Xue Y, et al. The necessity for standardization of glioma stem cell culture: a systematic review. Stem Cell Research & Therapy. 2020 Feb 26;11(1):84.

94. Gómez-Oliva R, Domínguez-García S, Carrascal L, Abalos-Martínez J, Pardillo-Díaz R, Verástegui C, et al. Evolution of Experimental Models in the Study of Glioblastoma: Toward Finding Efficient Treatments. Frontiers in Oncology. 2021;10.

95. Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, Donin NM, et al. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. Cancer Cell. 2006 May;9(5):391–403.

96. Galli R, Binda E, Orfanelli U, Cipelletti B, Gritti A, De Vitis S, et al. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. Cancer Res. 2004 Oct 1;64(19):7011–21.

97. Pollard SM, Yoshikawa K, Clarke ID, Danovi D, Stricker S, Russell R, et al. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. Cell Stem Cell. 2009 Jun 5;4(6):568–80.

98. Chen J, McKay RM, Parada LF. Malignant glioma: lessons from genomics, mouse models, and stem cells. Cell. 2012 Mar 30;149(1):36–47.

99. Wang J, Wakeman TP, Lathia JD, Hjelmeland AB, Wang XF, White RR, et al. Notch promotes radioresistance of glioma stem cells. Stem Cells. 2010 Jan;28(1):17–28.

100. Marbacher S, Fandino J, Kitchen ND. Standard intracranial in vivo animal models of delayed cerebral vasospasm. British Journal of Neurosurgery. 2010 Aug 1;24(4):415–34.

101. Abraham MK, Chang WTW. Subarachnoid Hemorrhage. Emergency Medicine Clinics of North America. 2016 Nov 1;34(4):901–16.

102. van Gijn J, Rinkel GJE. Subarachnoid haemorrhage: diagnosis, causes and management. Brain. 2001 Feb 1;124(2):249–78.

103. Macdonald RL, Schweizer TA. Spontaneous subarachnoid haemorrhage. The Lancet. 2017 Feb 11;389(10069):655–66.

104. de Rooij NK, Linn FHH, van der Plas JA, Algra A, Rinkel GJE. Incidence of subarachnoid haemorrhage: a systematic review with emphasis on region, age, gender and time trends. J Neurol Neurosurg Psychiatry. 2007 Dec;78(12):1365–72.

105. Linn F h. h., Rinkel G j. e., Algra A, van Gijn J. Incidence of Subarachnoid Hemorrhage. Stroke. 1996 Apr;27(4):625–9.

106. Chan V, Lindsay P, McQuiggan J, Zagorski B, Hill MD, O'Kelly C. Declining Admission and Mortality Rates for Subarachnoid Hemorrhage in Canada Between 2004 and 2015. Stroke. 2019 Jan;50(1):181–4.

107. Etminan N, Chang HS, Hackenberg K, de Rooij NK, Vergouwen MDI, Rinkel GJE, et al. Worldwide Incidence of Aneurysmal Subarachnoid Hemorrhage According to Region, Time Period, Blood Pressure, and Smoking Prevalence in the Population. JAMA Neurol. 2019 May;76(5):588–97.

108. Giordan E, Graffeo CS, Rabinstein AA, Brown RD, Rocca WA, Chamberlain AM, et al. Aneurysmal subarachnoid hemorrhage: long-term trends in incidence and survival in Olmsted County, Minnesota. Journal of Neurosurgery. 2020 Feb 21;134(3):878–83.

109. Okhovat AA. Chapter Six(c) - Thunderclap headache. In: Togha M, Jafari E, Mohammadianinejad SE, Haghighi S, Ansari H, editors. Headache and Migraine in Practice [Internet]. Academic Press; 2022 [cited 2024 Jun 4]. p. 141–50. Available from: https://www.sciencedirect.com/science/article/pii/B9780323997294000235

110. Marx J, Hockberger R, Walls R. Rosen's Emergency Medicine - Concepts and Clinical Practice E-Book: 2-Volume Set. Elsevier Health Sciences; 2013. 3266 p.

111. Green DM, Burns JD, DeFusco CM. ICU Management of Aneurysmal Subarachnoid Hemorrhage. J Intensive Care Med. 2013 Nov 1;28(6):341–54.

112. Edlow JA, Fisher J. Diagnosis of Subarachnoid Hemorrhage. Stroke. 2012 Aug;43(8):2031–2.

113. Deutsche Gesellschaft für Neurologie e.V. (DGN). Leitlinie: Subarachnoidalblutung [Internet]. 2012 [cited 2024 May 31]. Available from: https://dgn.org/leitlinie/subarachnoidalblutung-sab

114. Fisher CM, Kistler JP, Davis JM. Relation of Cerebral Vasospasm to Subarachnoid Hemorrhage Visualized by Computerized Tomographic Scanning. Neurosurgery. 1980 Jan;6(1):1.

115. Frontera JA, Claassen J, Schmidt JM, Wartenberg KE, Temes R, Connolly ES, et al. PREDICTION OF SYMPTOMATIC VASOSPASMAFTER SUBARACHNOID HEMORRHAGE: THE MODIFIED FISHER SCALE. Neurosurgery. 2006 Jul;59(1):21.

116. Rouanet C, Silva GS. Aneurysmal subarachnoid hemorrhage: current concepts and updates. Arq Neuro-Psiquiatr. 2019 Dec 5;77:806–14.

117. Osgood ML. Aneurysmal Subarachnoid Hemorrhage: Review of the Pathophysiology and Management Strategies. Curr Neurol Neurosci Rep. 2021 Jul 26;21(9):50.

118. Tang C, Zhang TS, Zhou LF. Risk Factors for Rebleeding of Aneurysmal Subarachnoid Hemorrhage: A Meta-Analysis. PLoS One. 2014 Jun 9;9(6):e99536.

119. Lanzino G, D'Urso PI, Suarez J, The Participants in the International Multi-disciplinary Consensus Conference on the Critical Care Management of Subarachnoid Hemorrhage. Seizures and Anticonvulsants after Aneurysmal Subarachnoid Hemorrhage. Neurocrit Care. 2011 Jul 13;15(2):247.

120. Kusaka G, Ishikawa M, Nanda A, Granger DN, Zhang JH. Signaling Pathways for Early Brain Injury after Subarachnoid Hemorrhage. J Cereb Blood Flow Metab. 2004 Aug 1;24(8):916–25.

121. Suzuki H. What is Early Brain Injury? Transl Stroke Res. 2015 Feb 1;6(1):1–3.

122. Rass V, Helbok R. Early Brain Injury After Poor-Grade Subarachnoid Hemorrhage. Curr Neurol Neurosci Rep. 2019;19(10):78.

123. Gaasch M, Schiefecker AJ, Kofler M, Beer R, Rass V, Pfausler B, et al. Cerebral Autoregulation in the Prediction of Delayed Cerebral Ischemia and Clinical Outcome in Poor-Grade Aneurysmal Subarachnoid Hemorrhage Patients*. Critical Care Medicine. 2018 May;46(5):774.

124. Hayman EG, Wessell A, Gerzanich V, Sheth KN, Simard JM. Mechanisms of global cerebral edema formation in aneurysmal subarachnoid hemorrhage. Neurocrit Care. 2017 Apr;26(2):301–10.

125. Weimer JM, Jones SE, Frontera JA. Acute Cytotoxic and Vasogenic Edema after Subarachnoid Hemorrhage: A Quantitative MRI Study. AJNR Am J Neuroradiol. 2017 May;38(5):928–34.

126. Fujii M, Yan J, Rolland WB, Soejima Y, Caner B, Zhang JH. Early brain injury, an evolving frontier in subarachnoid hemorrhage research. Transl Stroke Res. 2013 Aug;4(4):432–46.

127. Geraghty JR, Testai FD. Delayed Cerebral Ischemia after Subarachnoid Hemorrhage: Beyond Vasospasm and Towards a Multifactorial Pathophysiology. Curr Atheroscler Rep. 2017 Oct 23;19(12):50.

128. Chen S, Li Q, Wu H, Krafft PR, Wang Z, Zhang JH. The harmful effects of subarachnoid hemorrhage on extracerebral organs. BioMed Research International. 2014;2014.

129. Øie LR, Solheim O, Majewska P, Nordseth T, Müller TB, Carlsen SM, et al. Incidence and case fatality of aneurysmal subarachnoid hemorrhage admitted to hospital between 2008 and 2014 in Norway. Acta Neurochir. 2020 Sep 1;162(9):2251–9.

130. Stegmayr B, Eriksson M, Asplund K. Declining Mortality From Subarachnoid Hemorrhage. Stroke. 2004 Sep;35(9):2059–63.

131. Worthington JM, Goumas C, Jalaludin B, Gattellari M. Decreasing Risk of Fatal Subarachnoid Hemorrhage and Other Epidemiological Trends in the Era of Coiling Implementation in Australia. Frontiers in Neurology. 2017;8.

132. Chu KH, Mahmoud I, Hou XY, Winter CD, Jeffree RL, Brown NJ, et al. Incidence and outcome of subarachnoid haemorrhage in the general and emergency department populations in Queensland from 2010 to 2014. Emergency Medicine Australasia. 2018;30(4):503–10.

133. Rinkel GJ, Algra A. Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage. The Lancet Neurology. 2011 Apr 1;10(4):349–56.

134. Johnston SC, Selvin S, Gress DR. The burden, trends, and demographics of mortality from subarachnoid hemorrhage. Neurology. 1998;50(5):1413–8.

135. Barry KJ, Gogjian MA, Stein BM. Small animal model for investigation of subarachnoid hemorrhage and cerebral vasospasm. Stroke. 1979 Sep 1;10(5):538–41.

136. Buehler D, Schueller K, Plesnila N. Protocol for the Induction of Subarachnoid Hemorrhage in Mice by Perforation of the Circle of Willis with an Endovascular Filament. Transl Stroke Res. 2014 Dec;5(6):653–9.

137. Sehba FA. Rat endovascular perforation model. Transl Stroke Res. 2014 Dec;5(6):660–8.

138. Thal SC, Plesnila N. Non-invasive intraoperative monitoring of blood pressure and arterial pCO2 during surgical anesthesia in mice. Journal of Neuroscience Methods. 2007 Jan 30;159(2):261–7.

139. Yanagisawa T, Zhang H, Suzuki T, Kamio Y, Takizawa T, Morais A, et al. Sex and Genetic Background Effects on the Outcome of Experimental Intracranial Aneurysms. Stroke. 2020 Oct;51(10):3083–94.

140. Weyer V, Maros ME, Kirschner S, Krost-Reuhl S, Groden C, Kramer M, et al. Influence of neurovascular anatomy on perforation site in different mouse strains using the filament perforation model for induction of subarachnoid hemorrhage. PLoS One. 2022 Oct 13;17(10):e0263983.

141. Dalrymple-Alford JC, Benton D. Preoperative differential housing and dorsal hippocampal lesions in rats. Behav Neurosci. 1984 Feb;98(1):23–34.

142. E. U. Directive. 63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Off J Eur Union. 2010;276:33–79.

143. Baumans V, Van Loo PLP. How to improve housing conditions of laboratory animals: the possibilities of environmental refinement. Vet J. 2013 Jan;195(1):24–32.

144. Grimm MS, Emerman JT, Weinberg J. Effects of social housing condition and behavior on growth of the Shionogi mouse mammary carcinoma. Physiol Behav. 1996;59(4–5):633–42.

145. Pham TM, Hagman B, Codita A, Van Loo PLP, Strömmer L, Baumans V. Housing environment influences the need for pain relief during post-operative recovery in mice. Physiol Behav. 2010 Apr 19;99(5):663–8.

146. Weber EM, Dallaire JA, Gaskill BN, Pritchett-Corning KR, Garner JP. Aggression in group-housed laboratory mice: why can't we solve the problem? Lab Anim (NY). 2017 Mar 22;46(4):157–61.

147. Lidington D., Wan H., Dinh D.D., Ng C., Bolz S.-S. Circadian Rhythmicity in Cerebral Microvascular Tone Influences Subarachnoid Hemorrhage-Induced Injury. Stroke. 2022;53(1):249–59.

148. Verweij M, van Ginhoven TM, Mitchell JR, Sluiter W, den Engel S van, Roest HP, et al. Preoperative fasting protects mice against hepatic ischemia/reperfusion injury: mechanisms and effects on liver regeneration. Liver Transplantation. 2011;17(6):695–704.

149. Gisselsson L, Smith ML, Siesjö BK. Hyperglycemia and Focal Brain Ischemia. J Cereb Blood Flow Metab. 1999 Mar 1;19(3):288–97.

150. Lewis SR, Evans DJ, Butler AR, Schofield-Robinson OJ, Alderson P. Hypothermia for traumatic brain injury. Cochrane Database of Systematic Reviews. 2017;(9).

151. Nicolaus ML, Bergdall VK, Davis IC, Hickman-Davis JM. Effect of Ventilated Caging on Water Intake and Loss in 4 Strains of Laboratory Mice. J Am Assoc Lab Anim Sci. 2016 Sep;55(5):525–33.

152. Tang X, Orchard SM, Sanford LD. Home cage activity and behavioral performance in inbred and hybrid mice. Behav Brain Res. 2002 Nov 15;136(2):555–69.

153. Björkman S, Nilsson F, Akeson J, Messeter K, Rosén I. The effect of thiopental on cerebral blood flow, and its relation to plasma concentration, during simulated induction of anaesthesia in a porcine model. Acta Anaesthesiol Scand. 1994 Jul;38(5):473–8.

154. Oshima T, Karasawa F, Satoh T. Effects of propofol on cerebral blood flow and the metabolic rate of oxygen in humans. Acta Anaesthesiol Scand. 2002 Aug;46(7):831–5.

155. Kamii H, Kato I, Kinouchi H, Chan PH, Epstein CJ, Akabane A, et al. Amelioration of vasospasm after subarachnoid hemorrhage in transgenic mice overexpressing CuZn-superoxide dismutase. Stroke. 1999 Apr;30(4):867–71.

156. 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. Frontiers in Molecular Neuroscience. 2018;11.

157. Feiler S, Friedrich B, Schöller K, Thal SC, Plesnila N. Standardized induction of subarachnoid hemorrhage in mice by intracranial pressure monitoring. J Neurosci Methods. 2010 Jul 15;190(2):164–70.

158. Bühler D, Schüller K, Plesnila N. Protocol for the induction of subarachnoid hemorrhage in mice by perforation of the Circle of Willis with an endovascular filament. Transl Stroke Res. 2014 Dec;5(6):653–9.

159. Adamson TW, Kendall LV, Goss S, Grayson K, Touma C, Palme R, et al. Assessment of Carprofen and Buprenorphine on Recovery of Mice after Surgical Removal of the Mammary Fat Pad. J Am Assoc Lab Anim Sci. 2010 Sep;49(5):610–6.

160. Arras M, Becker K, Bergadano A, Durst M, Eberspächer-Schweda E, Fleischmann T, et al. Fachinformation aus dem Ausschuss für Anästhesie und Analgesie der GV-SOLAS in Zusammenarbeit mit dem Arbeitskreis 4 in der TVT - Schmerztherapie bei Versuchstieren [Internet]. 2020 [cited 2024 Jun 1]. Available from: https://www.gv-solas.de/wp-content/uploads/2020/09/Schmerztherapie-bei-Versuchstieren_2020.pdf

161. Staib-Lasarzik I, Nagel N, Sebastiani A, Griemert EV, Thal SC. Analgesic treatment limits surrogate parameters for early stress and pain response after experimental subarachnoid hemorrhage. BMC Neurosci. 2019 Sep 18;20(1):49.

162. Sehba FA, Pluta RM. Aneurysmal subarachnoid hemorrhage models: do they need a fix? Stroke Res Treat. 2013;2013:615154.

163. Prunell GF, Mathiesen T, Svendgaard NA. A new experimental model in rats for study of the pathophysiology of subarachnoid hemorrhage. Neuroreport. 2002 Dec 20;13(18):2553–6.

164. Leonardo C, Robbins S, Dore S. Translating Basic Science Research to Clinical Application: Models and Strategies for Intracerebral Hemorrhage. Frontiers in Neurology. 2012;3.

165. Gules I, Satoh M, Clower BR, Nanda A, Zhang JH. Comparison of three rat models of cerebral vasospasm. American Journal of Physiology-Heart and Circulatory Physiology. 2002 Dec;283(6):H2551–9.

166. Russel WMS, Burch RL. The Principles of Humane Experimental Technique. Medical Journal of Australia. 1960;1(13):500–500.

167. Sugawara T, Ayer R, Jadhav V, Zhang JH. A new grading system evaluating bleeding scale in filament perforation subarachnoid hemorrhage rat model. J Neurosci Methods. 2008 Jan 30;167(2):327–34.

168. Egashira Y, Shishido H, Hua Y, Keep RF, Xi G. New grading system based on magnetic resonance imaging in a mouse model of subarachnoid hemorrhage. Stroke. 2015 Feb;46(2):582–4.

169. Mutoh T, Mutoh T, Sasaki K, Nakamura K, Taki Y, Ishikawa T. Value of Three-Dimensional Maximum Intensity Projection Display to Assist in Magnetic Resonance Imaging (MRI)-Based Grading in a Mouse Model of Subarachnoid Hemorrhage. Med Sci Monit. 2016 Jun 16;22:2050–5.

170. Shuheng Liu, Katharina Tielking, Dario von Wedel, Melina Nieminen-Kelhä, Susanne Mueller, Philipp Boehm-Sturm, et al. Endovascular Perforation Model for Subarachnoid Hemorrhage Combined with Magnetic Resonance Imaging (MRI). J Vis Exp. 2021;178.

171. Lin C.-L., Calisaneller T., Ukita N., Dumont A.S., Kassell N.F., Lee K.S. A murine model of subarachnoid hemorrhage-induced cerebral vasospasm. J Neurosci Methods. 2003;123(1):89–97.

172. Kassell NF, Torner JC, Haley EC, Jane JA, Adams HP, Kongable GL. The International Cooperative Study on the Timing of Aneurysm Surgery. Part 1: Overall management results. J Neurosurg. 1990 Jul;73(1):18–36.

173. Lord AS, Fernandez L, Schmidt JM, Mayer SA, Claassen J, Lee K, et al. Effect of rebleeding on the course and incidence of vasospasm after subarachnoid hemorrhage. Neurology. 2012 Jan 3;78(1):31–7.

174. Percie du 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 Jul 14;18(7):e3000411.

175. Lazic SE, Clarke-Williams CJ, Munafò MR. What exactly is 'N' in cell culture and animal experiments? PLOS Biology. 2018 Apr 4;16(4):e2005282.

176. Tsvetkov D, Kolpakov E, Kassmann M, Schubert R, Gollasch M. Distinguishing Between Biological and Technical Replicates in Hypertension Research on Isolated Arteries. Frontiers in Medicine. 2019;6.

177. Emmerich C. Accurate design of in vitro experiments – why does it matter? [Internet]. Partnership for Assessment and Accreditation of Scientific Practice (paasp) network. 2016 [cited 2024 Jun 1]. Available from: https://paasp.net/accurate-design-of-in-vitro-experiments-why-does-it-matter/

178. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J (Clin Res Ed). 1986 Mar 15;292(6522):746–50.

179. US Pharmacopeia. USP Monographs: Nonabsorbable Surgical Suture [Internet]. 2023 [cited 2024 Jun 1]. Available from: http://ftp.uspbpep.com/v29240/usp29nf24s0_m80200.html

180. Wisdom JP, Cavaleri MA, Onwuegbuzie AJ, Green CA. Methodological Reporting in Qualitative, Quantitative, and Mixed Methods Health Services Research Articles. Health Services Research. 2012;47(2):721–45.

181. NEJM Group. NEJM Evidence: Author Center [Internet]. 2023 [cited 2024 Jun 1]. Available from: https://evidence.nejm.org/author-center/article-types-and-submission-information

182. The Lancet. The Lancet: Information for Authors [Internet]. 2023 [cited 2024 Jun 1]. Available from: https://www.thelancet.com/pb-assets/Lancet/authors/tl-info-for-authors-1676565160037.pdf

183. Kawakita F, Fujimoto M, Liu L, Nakano F, Nakatsuka Y, Suzuki H. Effects of Toll-Like Receptor 4 Antagonists Against Cerebral Vasospasm After Experimental Subarachnoid Hemorrhage in Mice. Mol Neurobiol. 2017 Oct;54(8):6624–33.

184. Hill A. The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. J Physiol. 1910;40((Suppl): iv–vii.).

185. Safdie F, Brandhorst S, Wei M, Wang W, Lee C, Hwang S, et al. Fasting Enhances the Response of Glioma to Chemo- and Radiotherapy. PLOS ONE. 2012 Sep 11;7(9):e44603.

186. Tieu MT, Lovblom LE, McNamara MG, Mason W, Laperriere N, Millar BA, et al. Impact of glycemia on survival of glioblastoma patients treated with radiation and temozolomide. J Neurooncol. 2015;124(1):119–26.

187. Lantigua H, Ortega-Gutierrez S, Schmidt JM, Lee K, Badjatia N, Agarwal S, et al. Subarachnoid hemorrhage: who dies, and why? Crit Care. 2015;19(1):309.

188. Park IS, Meno JR, Witt CE, Suttle TK, Chowdhary A, Nguyen TS, et al. Subarachnoid hemorrhage model in the rat: Modification of the endovascular filament model. Journal of Neuroscience Methods. 2008 Jul 30;172(2):195–200.

189. Trott AT. Wounds and Lacerations - Chapter 8 - Instruments and Suture Materials. In: Wounds and Lacerations (Third Edition). 3rd ed. Philadelphia: Mosby; 2005. p. 93–106.

190. DemeTECH. DemeLON MULTI ™ [Internet]. 2024 [cited 2024 Jun 1]. Available from: https://www.demetech.us/product/sutures/NylonMultifilament

191. Vojgani Y, Shirazi A, Zarei S, Yeganeh O, Jeddi-Tehrani M. Comparison of efficacies of fetal bovine sera from different suppliers in cell culture experiments. Comp Clin Pathol. 2018 Mar 1;27(2):519–27.

192. Ryu AH, Eckalbar WL, Kreimer A, Yosef N, Ahituv N. Use antibiotics in cell culture with caution: genome-wide identification of antibiotic-induced changes in gene expression and regulation. Sci Rep. 2017 Aug 8;7:7533.

193. Animal Resource Program (Pennsylvania State University). Research at Penn State. 2018 [cited 2023 May 29]. Anesthesia in Experimental Animals. Available from: https://www.research.psu.edu/animalresourceprogram/surgery/anesthesia-rodents-rabbits

194. Priebe HJ. Aneurysmal subarachnoid haemorrhage and the anaesthetist. BJA: British Journal of Anaesthesia. 2007 Jul 1;99(1):102–18.

195. Suman PR, Oliveira CL de. Systematic heterogenisation to improve reproducibility in animal studies. PLOS Biology. 2022 May 6;20(5):e3001629.

196. Wang L, Jung J, Babikir H, Shamardani K, Jain S, Feng X, et al. A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. Nat Cancer. 2022 Dec;3(12):1534–52.

197. Wang MD, Fu QH, Song MJ, Ma WB, Zhang JH, Wang ZX. Novel Subgroups in Subarachnoid Hemorrhage and Their Association With Outcomes—A Systematic Review and Meta-Regression. Frontiers in Aging Neuroscience. 2021;12.

198. Sottoriva A, Spiteri I, Piccirillo SGM, Touloumis A, Collins VP, Marioni JC, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proceedings of the National Academy of Sciences. 2013 Mar 5;110(10):4009–14.

199. D'Souza S. Aneurysmal Subarachnoid Hemorrhage. J Neurosurg Anesthesiol. 2015 Jul;27(3):222–40.

200. Hirst TC, Vesterinen HM, Sena ES, Egan KJ, Macleod MR, Whittle IR. Systematic review and meta-analysis of temozolomide in animal models of glioma: was clinical efficacy predicted? Br J Cancer. 2013 Jan;108(1):64–71.

201. Kamp, Lieshout JH van, Dibué-Adjei M, Weber JK, Schneider T, Restin T, et al. A Systematic and Meta-Analysis of Mortality in Experimental Mouse Models Analyzing Delayed Cerebral Ischemia After Subarachnoid Hemorrhage. Transl Stroke Res. 2017 Jun 1;8(3):206–19.

202. 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. J Cereb Blood Flow Metab. 2012 Sep 1;32(9):1645–58.

203. Dirnagl U. Bench to Bedside: The Quest for Quality in Experimental Stroke Research. J Cereb Blood Flow Metab. 2006 Dec 1;26(12):1465–78.

204. Sena ES, Worp HB van der, Bath PMW, Howells DW, Macleod MR. Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy. PLOS Biology. 2010 Mar 30;8(3):e1000344.

205. 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 Oct 1;39(10):2824–9.

206. Lalu MM, Montroy J, Begley CG, Bubela T, Hunniford V, Ripsman D, et al. Identifying and understanding factors that affect the translation of therapies from the laboratory to patients: a study protocol. F1000Res. 2020 Oct 16;9:485.

207. Caragher S, Chalmers AJ, Gomez-Roman N. Glioblastoma's Next Top Model: Novel Culture Systems for Brain Cancer Radiotherapy Research. Cancers (Basel). 2019 Jan 4;11(1):44.

208. De Vleeschouwer S, Bergers G. Glioblastoma: To Target the Tumor Cell or the Microenvironment? In: De Vleeschouwer S, editor. Glioblastoma. Brisbane (AU): Codon Publications; 2017.

209. Cras TY, Bos D, Ikram MA, Vergouwen MDI, Dippel DWJ, Voortman T, et al. Determinants of the Presence and Size of Intracranial Aneurysms in the General Population. Stroke. 2020 Jul;51(7):2103–10.

210. Turan N, Heider RAJ, Zaharieva D, Ahmad FU, Barrow DL, Pradilla G. Sex Differences in the Formation of Intracranial Aneurysms and Incidence and Outcome of Subarachnoid Hemorrhage: Review of Experimental and Human Studies. Transl Stroke Res. 2016 Feb 1;7(1):12–9.

211. Ahnstedt H, McCullough LD, Cipolla MJ. The Importance of Considering Sex Differences in Translational Stroke Research. Transl Stroke Res. 2016 Aug 1;7(4):261–73.

212. Ergul A, Hafez S, Fouda A, Fagan SC. Impact of Comorbidities on Acute Injury and Recovery in Preclinical Stroke Research: Focus on Hypertension and Diabetes. Transl Stroke Res. 2016 Aug 1;7(4):248–60.

213. Gröger M. Rolle der Kinin Rezeptoren beim sekundären Hirnschaden nach transienter fokaler zerebraler Ischämie der Maus. Medizinische Fakultät der Ludwig-Maximilians-Universität zu München; 2007.

214. Schröder H, Moser N, Huggenberger S. The Mouse Circle of Willis. In: Schröder H, Moser N, Huggenberger S, editors. Neuroanatomy of the Mouse: An Introduction. Cham: Springer International Publishing; 2020. p. 333–40.

215. Ikram A, Javaid MA, Ortega-Gutierrez S, Selim M, Kelangi S, Anwar SMH, et al. Delayed Cerebral Ischemia after Subarachnoid Hemorrhage. Journal of Stroke and Cerebrovascular Diseases. 2021 Nov 1;30(11):106064.

216. Beier D, Schriefer B, Brawanski K, Hau P, Weis J, Schulz JB, et al. Efficacy of clinically relevant temozolomide dosing schemes in glioblastoma cancer stem cell lines. J Neurooncol. 2012 Aug;109(1):45–52.

217. Stepanenko AA, Chekhonin VP. On the Critical Issues in Temozolomide Research in Glioblastoma: Clinically Relevant Concentrations and MGMT-independent Resistance. Biomedicines. 2019 Dec;7(4):92.

218. Strobel H, Baisch T, Fitzel R, Schilberg K, Siegelin MD, Karpel-Massler G, et al. Temozolomide and Other Alkylating Agents in Glioblastoma Therapy. Biomedicines. 2019 Sep 9;7(3):69.

219. Herbener VJ, Burster T, Goreth A, Pruss M, von Bandemer H, Baisch T, et al. Considering the Experimental Use of Temozolomide in Glioblastoma Research. Biomedicines. 2020 Jun;8(6):151.

220. ATCC, PBCF & PS-OC Bioresource Core Facility. SOP: Thawing, Propagation and Cryopreservation of NCI-PBCF-HTB14 (U-87 MG). 2012;24.

221. Oraiopoulou ME, Tzamali E, Tzedakis G, Vakis A, Papamatheakis J, Sakkalis V. In Vitro/In Silico Study on the Role of Doubling Time Heterogeneity among Primary Glioblastoma Cell Lines. Biomed Res Int. 2017;2017:8569328.

222. Stupp R, Brada M, van den Bent MJ, Tonn JC, Pentheroudakis G, ESMO Guidelines Working Group. High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2014 Sep;25 Suppl 3:iii93-101.

223. Muroi C, Fujioka M, Marbacher S, Fandino J, Keller E, Iwasaki K, et al. Mouse model of subarachnoid hemorrhage: technical note on the filament perforation model. Acta Neurochir Suppl. 2015;120:315–20.

224. Eastman A. Improving anticancer drug development begins with cell culture: misinformation perpetrated by the misuse of cytotoxicity assays. Oncotarget. 2016 Oct 14;8(5):8854–66.

225. Riss TL, Moravec RA, Niles AL, Duellman S, Benink HA, Worzella TJ, et al. Cell Viability Assays. In: Markossian S, Grossman A, Brimacombe K, Arkin M, Auld D, Austin C, et al., editors. Assay Guidance Manual. Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004.

226. Cai J, Sun Y, Yuan F, Chen L, He C, Bao Y, et al. A Novel Intravital Method to Evaluate Cerebral Vasospasm in Rat Models of Subarachnoid Hemorrhage: A Study with Synchrotron Radiation Angiography. PLOS ONE. 2012 Mar 12;7(3):e33366.

227. Buch K, Peters T, Nawroth T, Sänger M, Schmidberger H, Langguth P. Determination of cell survival after irradiation via clonogenic assay versus multiple MTT Assay - A comparative study. Radiation Oncology. 2012 Jan 3;7(1):1.

228. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.4 [Internet]. Cochrane; 2023 [cited 2024 Jun 1]. Available from: https://training.cochrane.org/handbook/current

229. Paez A. Gray literature: An important resource in systematic reviews. J Evid Based Med. 2017 Aug;10(3):233–40.

230. Page MJ, Sterne JAC, Higgins JPT, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. Research Synthesis Methods. 2021;12(2):248–59.

231. Borenstein M. Introduction to Meta-Analysis. John Wiley & Sons, Ltd; 2009.

232. Kadlec D, Sainani KL, Nimphius S. With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research. Sports Med. 2023 Feb 1;53(2):313–25.

233. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011 Feb 10;342:d549.

234. Hedges LV. The Statistics of Replication. Methodology. 2019 Oct;15(Supplement 1):3–14.

235. Noble WS. How does multiple testing correction work? Nat Biotechnol. 2009 Dec;27(12):1135–7.

236. Sander T, Wilson E, Ghanawi J, Macleod MR, Kahlert U. Study protocol: Systematic Review and Meta-Analysis of reporting on quality control standards and phenotypes of the basic cell model in brain cancer research: how reproducible is the field? [Internet]. OSF; 2020 [cited 2024 Jun 1]. Available from: https://osf.io/u6yhb/

237. Alpdogan S, Sander T, Zhang R, Muhammad S. Study protocol: Meta-review on SAH perforation model in mice [Internet]. Open Science Framework; 2021 [cited 2024 Jun 1]. Available from: https://osf.io/https://osf.io/sutfb

238. eLife sciences. eLife. eLife Sciences Publications Limited; 2014 [cited 2021 Apr 16]. Reproducibility Project: Cancer Biology. Available from: https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology

239. Hartshorne JK, Schachner A. Tracking Replicability as a Method of Post-Publication Open Evaluation. Front Comput Neurosci. 2012 Mar 5;6.

240. Roper K, Abdel-Rehim A, Hubbard S, Carpenter M, Rzhetsky A, Soldatova L, et al. Testing the reproducibility and robustness of the cancer biology literature by robot. Journal of The Royal Society Interface. 2022 Apr 6;19(189):20210821.

241. Wilson E, Cruz F, Maclean D, Ghanawi J, McCann S, Brennan P, et al. Screening for in vitro systematic reviews: a comparison of screening methods and training of a machine learning classifier. Clinical science (London, England : 1979). 2023 Jan 11;137.

242. King RD, Rowland J, Oliver SG, Young M, Aubrey W, Byrne E, et al. The automation of science. Science. 2009 Apr 3;324(5923):85–9.

243. Emmerich CH, Gamboa LM, Hofmann MCJ, Bonin-Andresen M, Arbach O, Schendel P, et al. Improving target assessment in biomedical research: the GOT-IT recommendations. Nat Rev Drug Discov. 2020 Nov 16;1–18.

244. The NPQIP Collaborative group. Did a change in Nature journals' editorial policy for life sciences research improve reporting? BMJ Open Science. 2019 Feb 1;3(1):e000035.

245. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Medical Research Methodology. 2014 Mar 26;14(1):43.

246. Laajala TD, Jumppanen M, Huhtaniemi R, Fey V, Kaur A, Knuuttila M, et al. Optimized design and analysis of preclinical intervention studies in vivo. Sci Rep. 2016 Aug 2;6:30723.

247. Krithikadatta J, Gopikrishna V, Datta M. CRIS Guidelines (Checklist for Reporting In-vitro Studies): A concept note on the need for standardized guidelines for improving quality and transparency in reporting in-vitro studies in experimental dental research. J Conserv Dent. 2014;17(4):301–4.

248. Ramirez FD, Jung RG, Motazedian P, Perry-Nguyen D, Di Santo P, MacDonald Z, et al. Journal Initiatives to Enhance Preclinical Research: Analyses of Stroke, Nature Medicine, Science Translational Medicine. Stroke. 2020 Jan;51(1):291–9.

249. Avey MT, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, et al. The Devil Is in the Details: Incomplete Reporting in Preclinical Animal Research. PLOS ONE. 2016 Nov 17;11(11):e0166733.

250. 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 May 24;13(5):e0197882.

251. Nature Communications. Nature: Nature Communications: For Authors: Article [Internet]. 2023 [cited 2024 Jun 1]. Available from: https://www.nature.com/ncomms/submit/article

252. PLOS. PLOS one: Submission guidelines [Internet]. 2023 [cited 2024 Jun 1]. Available from: https://journals.plos.org/plosone/s/submission-guidelines

253. Macleod, Mohan. Reproducibility and Rigor in Animal-Based Research. ILAR Journal. 2019 Dec 31;60(1):17–23.

254. Pedro-Roig L, Emmerich CH. The reproducibility crisis in preclinical research – lessons to learn from clinical research. MEW. 2017 Dec 1;26:28–32.

255. Samsa G, Samsa L. A Guide to Reproducibility in Preclinical Research. Acad Med. 2019 Jan;94(1):47–52.

256. Center for Open Science. Open Science Framework (OSF) [Internet]. 2023 [cited 2024 Jun 1]. Available from: https://osf.io/

257. CERN Data Centre & Invenio. Zenodo [Internet]. 2023 [cited 2024 Jun 3]. Available from: https://zenodo.org/

258. GitHub, Inc. GitHub. 2023 [cited 2023 Jul 3]. GitHub. Available from: https://github.com/

259. Cooper C, Bou JT, Varley-Campbell J. Evaluating the effectiveness, efficiency, cost and value of contacting study authors in a systematic review: a case study and worked example. BMC Medical Research Methodology. 2019 Mar 5;19(1):45.

260. Danko KJ, Dahabreh IJ, Ivers NM, Moher D, Grimshaw JM. Contacting authors by telephone increased response proportions compared with emailing: results of a randomized study. J Clin Epidemiol. 2019 Nov;115:150–9.

261. Voelkl B, Vogt L, Sena ES, Würbel H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. PLOS Biology. 2018 Feb 22;16(2):e2003693.

262. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution. 2013;4(2):133–42.

7 Supplementary material

7.1 Supplementary material to publication 1

The following supplementary material has been published alongside the abovementioned publication (5). The numeration of supplements has been maintained according to the original numeration used in the publication.

7.1.1 S1: Exact search strategy

(glioblastom* OR Astrocytom*, Grade IV OR Glioblastom* Multiform* OR Giant Cell Glioblastom* OR Brain cance* OR Malignant Gliom*) AND (Temozolomide OR Temodal OR Temodar OR methazolastone OR tmz) AND (U87 OR U-87 OR U87MG OR U87-MG OR U 87-MG OR U-87-MG OR U87 GBM OR U87-GBM OR U 87 GBM OR U-87 GBM OR U-87-GBM)

7.1.2 S2: Literature screening criteria

Inclusion criteria	Exclusion criteria
U-87 MG cell line as glioblastoma *in vitro* model	Other models than U-87 MG cell line *In vivo* models, Xenotransplantation models
TMZ single treatment	TMZ as a part of a combined treatment with other drugs or genetic interventions
Comparison of the effect of TMZ to an untreated control	No comparison to an untreated control
Cell viability assessment (MTT and similar colorimetric assays, cell counting) to quantify the effect of TMZ	Effect of TMZ measured with none of these cell viability assessment methods
DMEM as the cell culture medium	Other cell culture media than DMEM
Original peer-reviewed research articles	Other publication types (e.g., conference abstracts, poster presentations)
English language	Other languages than English

Articles were included if they met all inclusion criteria and no exclusion criteria. If an article included multiple experiments where one or more experiments did not match the criteria but at least one did match, then the article was included. DMEM = Dulbecco's Modified Eagle Medium; MTT = 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; TMZ = temozolomide; U-87 MG = Uppsala-87 Malignant Glioma.

7.1.3 S3: Extracted parameters from the included articles

Parameter	Possible phenotypes
General article information	
Title, Authors, Year, Publishing journals JIF	Not applicable
Conflicts of interest's statement	Conflicts / no conflicts
Cell model	
U-87 MG cell line source	Name and country of source
U-87 MG cell line authentication reporting	Reported / not reported
U-87 MG cell line age reporting	Reported / not reported
Cell culture conditions	
Number/Volume/Concentration of U-87 MG cells	Exact Number/Volume/Concentration
Cell passaging criterion	Confluency/Time intervals
Glucose level of cell culture medium	Exact glucose concentration / high glucose / low glucose / no glucose
Reporting of successful mycoplasma exclusion test	Reported / not reported
Supplemented antibiotics	Name and dose of the antibiotics
FBS use and source	FBS supplemented / FBS not supplemented Name and country of source
Intervention & control group	
TMZ source	Name and country of source
Volume of the added TMZ suspension	Exact volume
Type of untreated control	Drug vehicle / medium only / other
Volume of the added control suspension	Exact volume
Outcome	
Cell viability of U-87 MG cells after incubation with TMZ and the corresponding untreated control	Mean and error data of cell viabilities / growth inhibition rates / proliferation rates
Concentration of TMZ	Exact concentration
Treatment duration	Exact duration
Number of experiments	Exact number of experiments
Type of outcome measurement assay	Name of assay

FBS = Fetal bovine serum; JIF = Journal Impact Factor (for the year of publication; obtained from Clarivates InCites Journal Citation Reports); U-87 MG = Uppsala-87 Malignant Glioma; TMZ = Temozolomide.

7.1.4 S4: Assessed potential risks of bias parameters

Were data for every relevant experiment mentioned in the methods section of a particular article presented?
Were data for every concentration of temozolomide and treatment duration presented?
Were the number of experiments and the number of replicates per experiment clearly reported?
Was a sample size calculation for the needed number of experiments reported?
Was the way calculating the U87-MG cell viability mean and error data clearly reported?
Was it reported whether the authors allocated the U87-MG cells randomly to the treatment and control group?
Was it reported whether the authors measured the U87-MG cell viability blinded?
Was a pre-registered study protocol available?

U-87 MG = Uppsala-87 Malignant Glioma.

7.1.5 S5: Parameters the authors were asked for in case of non-reporting

Missing information in the articles:
U-87 MG cell line authentication
U-87 MG age (maximum number of passages)
Glucose level of cell culture medium
Type of untreated control
Concentration of TMZ
Number of experiments and replicates per experiment
Type of error of the presented data (SD or SEM)
Treatment duration
Additional risks of bias parameters:
Therapy regime (single dose, multi dose or different)
Was the cell viability measured directly after the treatment of U-87 MG cells with TMZ?
Way of calculation of cell viability data (mean of all data, mean of means or different)

SD = standard deviation; SEM = standard error of the mean; U-87 MG = Uppsala-87 Malignant Glioma; TMZ = Temozolomide.

7.1.6 S6: Full list of included articles into the systematic review and meta-analysis

Article number	Title	Authors	Year	DOI	Included in meta-analysis
1	Afatinib and Temozolomide combination inhibits tumorigenesis by targeting EGFRvIII-cMet signaling in glioblastoma cells	Vengoji et al.	2019	10.1186/s13046-019-1264-2	TRUE
2	Akt and beta-catenin contribute to TMZ resistance and EMT of MGMT negative malignant glioma cell line	Yi et al.	2016	10.1016/j.jns.2016.05.054	TRUE
3	Anti-tumor activities of luteolin and silibinin in glioblastoma cells: overexpression of miR-7-1-3p augmented luteolin and silibinin to inhibit autophagy and induce apoptosis in glioblastoma in vivo	Chakrabarti et al.	2015	10.1007/s10495-015-1198-x	TRUE
4	Anticancer activity of flavonoids isolated from Achyrocline satureioides in gliomas cell lines	De Souza et al.	2018	10.1016/j.tiv.2018.04.013	TRUE
5	Artesunate Enhances the Antiproliferative Effect of Temozolomide on U87MG and A172 Glioblastoma Cell Lines	Karpel-Massler et al.	2014	10.2174/187152061131366 60340	TRUE
6	ATM inhibitor KU-55933 increases the TMZ responsiveness of only inherently TMZ sensitive GBM cells	Nadkarni et al.	2012	10.1007/s11060-012-0979-0	TRUE
7	Berberine induces senescence of human glioblastoma cells by downregulating the EGFR-MEK-ERK signaling pathway	Liu et al.	2014	10.1158/1535-7163.MCT-14-0634	TRUE
8	beta-elemene enhances both radiosensitivity and chemosensitivity of glioblastoma cells through the inhibition of the ATM signaling pathway	Liu et al.	2015	10.3892/or.2015.4050	TRUE

9	Blocking LDHA glycolytic pathway sensitizes glioblastoma cells to radiation and temozolomide	Koukourakis et al.	2017	10.1016/j.bbrc.2017.07.138	TRUE
10	Bortezomib inhibits growth and sensitizes glioma to temozolomide (TMZ) via down-regulating the FOXM1-Survivin axis	Tang et al.	2019	10.1186/s40880-019-0424-2	TRUE
11	Bortezomib overcomes MGMT-related resistance of glioblastoma cell lines to temozolomide in a schedule-dependent manner	Vlachostergios et al.	2013	10.1007/s10637-013-9968-1	TRUE
12	Bufothionine Promotes Apoptosis via Triggering ER Stress and Synergizes with Temozolomide in Glioblastoma Multiforme Cells	Sun et al.	2019	10.1002/ar.24194	TRUE
13	Calpain suppresses cell growth and invasion of glioblastoma multiforme by producing the cleavage of filamin A	Cai et al.	2020	10.1007/s10147-020-01636-7	TRUE
14	Chemotherapeutic effect of tamoxifen on temozolomide-resistant gliomas	He et al.	2015	10.1097/CAD.0000000000000197	TRUE
15	Chloroquine enhances temozolomide cytotoxicity in malignant gliomas by blocking autophagy	Golden et al.	2014	10.3171/2014.9.FOCUS14504	TRUE
16	Chronic exposure of human glioblastoma tumors to low concentrations of a pesticide mixture induced multidrug resistance against chemotherapy agents	Doganlar et al.	2020	10.1016/j.ecoenv.2020.110940	TRUE
17	Combination of biochanin a and temozolomide impairs tumor growth by modulating cell metabolism in glioblastoma multiforme	Desai et al.	2019	10.21873/anticanres.13079	TRUE

18	Combination of caspase transfer using the human telomerase reverse transcriptase promoter and conventional therapies for malignant glioma cells	Takeuchi et al.	2004	10.3892/ijo.25.1.57	TRUE
19	Combination of the mTOR inhibitor RAD001 with temozolomide and radiation effectively inhibits the growth of glioblastoma cells in culture	Burckel et al.	2014	10.3892/or.2014.3590	TRUE
20	Combined effects of mesenchymal stem cells carrying cytosine deaminase gene with 5-fluorocytosine and temozolomide in orthotopic glioma model	Chang et al.	2020		TRUE
21	CtIP contributes to non-homologous end joining formation through interacting with ligase IV and promotion of TMZ resistance in glioma cells	Yang et al.	2019	10.26355/eurrev_201903_17252	TRUE
22	Cytotoxicity of temozolomide on human glioblastoma cells is enhanced by the concomitant exposure to an extremely low-frequency electromagnetic field (100 Hz 100 G)	Akbarnejad et al.	2017	10.1016/j.biopha.2017.05.050	TRUE
23	Development of transferrin-modified poly(lactic-co-glycolic acid) nanoparticles for glioma therapy	Mao et al.	2019	10.1097/CAD.0000000000000754	TRUE
24	Do Anti-Oxidants Vitamin D(3) Melatonin and Alpha-Lipoic Acid Have Synergistic Effects with Temozolomide on Cultured Glioblastoma Cells?	McConnell et al.	2018	10.3390/medicines5020058	TRUE
25	Down-Regulation of AQP4 Expression via p38 MAPK Signaling in Temozolomide-Induced Glioma Cells Growth Inhibition and Invasion Impairment	Chen et al.	2017	10.1002/jcb.26176	TRUE

26	Effect of the STAT3 inhibitor STX-0119 on the proliferation of a temozolomide-resistant glioblastoma cell line	Ashizawa et al.	2014	10.3892/ijo.2014.2439	TRUE
27	Effects of solvent used for fabrication on drug loading and release kinetics of electrosprayed temozolomide-loaded PLGA microparticles for the treatment of glioblastoma	Rodriguez de Anda et al.	2019	10.1002/jbm.b.34324	TRUE
28	Effects of temozolomide (TMZ) on the expression and interaction of heat shock proteins (HSPs) and DNA repair proteins in human malignant glioma cells	Castro et al.	2014	10.1007/s12192-014-0537-0	TRUE
29	EGCG inhibits properties of glioma stem-like cells and synergizes with temozolomide through downregulation of P-glycoprotein inhibition	Zhang et al.	2014	10.1007/s11060-014-1604-1	TRUE
30	EMAP-II sensitize U87MG and glioma stem-like cells to temozolomide via induction of autophagy-mediated cell death and G2/M arrest	Yu et al.	2017	10.1080/15384101.2017.1315492	TRUE
31	Enhancing glioblastoma cell sensitivity to chemotherapeutics: A strategy involving survivin gene silencing mediated by gemini surfactant-based complexes	Cruz et al.	2016	10.1016/j.ejpb.2016.04.014	TRUE
32	Exogenous IGFBP-2 promotes proliferation invasion and chemoresistance to temozolomide in glioma cells via the integrin beta 1-ERK pathway	Han et al.	2014	10.1038/bjc.2014.435	TRUE
33	Fever-Range Hyperthermia vs. Hypothermia Effect on Cancer Cell Viability Proliferation and HSP90 Expression	Kalamida et al.	2015	10.1371/journal.pone.0116021	TRUE

34	FTY720 inhibits the Nrf2/ARE pathway in human glioblastoma cell lines and sensitizes glioblastoma cells to temozolomide	Zhang et al.	2017	10.1016/j.pharep.2017.07.003	TRUE
35	G3BP1 knockdown sensitizes U87 glioblastoma cell line to Bortezomib by inhibiting stress granules assembly and potentializing apoptosis	Bittencourt et al.	2019	10.1007/s11060-019-03252-6	TRUE
36	GADD45A plays a protective role against temozolomide treatment in glioblastoma cells	Wang et al.	2017	10.1038/s41598-017-06851-3	TRUE
37	Gene expression profiling predicts response to temozolomide in malignant gliomas	Yoshino et al.	2010	10.3892/ijo_00000621	TRUE
38	Genomic profiling of long non-coding RNA and mRNA expression associated with acquired temozolomide resistance in glioblastoma cells	Zheng et al.	2017	10.3892/ijo.2017.4033	TRUE
39	Glucosylceramide synthase silencing combined with the receptor tyrosine kinase inhibitor axitinib as a new multimodal strategy for glioblastoma	Morais et al.	2019	10.1093/hmg/ddz152	TRUE
40	Growth Inhibitory Effects of Dipotassium Glycyrrhizinate in Glioblastoma Cell Lines by Targeting MicroRNAs Through the NF-kappa B Signaling Pathway	Bonafe et al.	2019	10.3892/ijmm.2015.2312	TRUE
41	Heterogeneous glioblastoma cell cross-talk promotes phenotype alterations and enhanced drug resistance	Motaln et al.	2015	10.18632/oncotarget.5701	TRUE
42	High-throughput screening uncovers miRNAs enhancing glioblastoma cell susceptibility to tyrosine kinase inhibitors	Cunha et al.	2017	10.1093/hmg/ddx323	TRUE

43	Honokiol enhances temozolomide-induced apoptotic insults to malignant glioma cells via an intrinsic mitochondrion-dependent pathway	Chio et al.	2018	10.1016/j.phymed.2018.06.012	TRUE
44	IDH1 R132H mutation regulates glioma chemosensitivity through Nrf2 pathway	Li et al.	2017	10.18632/oncotarget.15868	TRUE
45	Improved effects of honokiol on temozolomide-induced autophagy and apoptosis of drug-sensitive and -tolerant glioma cells	Chio et al.	2018	10.1186/s12885-018-4267-z	TRUE
46	In vitro and in vivo effect of human lactoferrin on glioblastoma growth	Arcella et al.	2015	10.3171/2014.12.JNS145112	TRUE
47	In vitro novel combinations of psychotropics and anti-cancer modalities in U87 human glioblastoma cells	Tzadok et al.	2010	10.3892/ijo-00000756	TRUE
48	In vitro radiosensitizing effects of temozolomide on U87MG cell lines of human glioblastoma multiforme	Borhani et al.	2017		TRUE
49	Induction of microRNA-146a is involved in curcumin-mediated enhancement of temozolomide cytotoxicity against human glioblastoma	Wu et al.	2015	10.3892/mmr.2015.4087	TRUE
50	Inhibition of STAT3 reverses alkylator resistance through modulation of the AKT and beta-catenin signaling pathways	Wang et al.	2011	10.3892/or.2011.1396	TRUE
51	Inhibition of telomerase activity in malignant glioma cells correlates with their sensitivity to temozolomide	Kanzawa et al.	2003	10.1038/sj.bjc.6601193	TRUE
52	Lithium enhances the antitumour effect of temozolomide against TP53 wild-type glioblastoma cells via NFAT1/FasL signalling	Han et al.	2017	10.1038/bjc.2017.89	TRUE

53	Magnolol and honokiol exert a synergistic anti-tumor effect through autophagy and apoptosis in human glioblastomas	Cheng et al.	2016	10.18632/oncotarget.8674	TRUE
54	Major Contribution of Caspase-9 to Honokiol-Induced Apoptotic Insults to Human Drug-Resistant Glioblastoma Cells	Wu et al.	2020	10.3390/molecules25061450	TRUE
55	Mechanisms and antitumor activity of a binary EGFR/DNA-targeting strategy overcomes resistance of glioblastoma stem cells to temozolomide	Sharifi et al.	2019	10.1158/1078-0432.CCR-19-0955	TRUE
56	MicroRNA-182 targets protein phosphatase 1 regulatory inhibitor subunit 1C in glioblastoma	Liu et al.	2017	10.18632/oncotarget.21309	TRUE
57	MicroRNA-21 silencing enhances the cytotoxic effect of the antiangiogenic drug sunitinib in glioblastoma	Costa et al.	2013	10.1093/hmg/dds496	TRUE
58	MicroRNA-29b promotes cell sensitivity to Temozolomide by targeting STAT3 in glioma	Xu et al.	2020	10.26355/eurrev_202002_20370	TRUE
59	MIM1 the Mcl-1 - specific BH3 mimetic induces apoptosis in human U87MG glioblastoma cells	Respondek et al.	2018	10.1016/j.tiv.2018.08.007	TRUE
60	MiR-144 overexpression as a promising therapeutic strategy to overcome glioblastoma cell invasiveness and resistance to chemotherapy	Cardoso et al.	2019	10.1093/hmg/ddz099	TRUE
61	miR-203 sensitizes glioma cells to temozolomide and inhibits glioma cell invasion by targeting E2F3	Tang et al.	2015	10.3892/mmr.2014.3101	TRUE

62	MiR-519a enhances chemosensitivity and promotes autophagy in glioblastoma by targeting STAT3/Bcl2 signaling pathway	Li et al.	2018	10.1186/s13045-018-0618-0	TRUE
63	Mitochondria Transcription Factor A: A Putative Target for the Effect of Melatonin on U87MG Malignant Glioma Cell Line	Franco et al.	2018	10.3390/molecules23051129	TRUE
64	N-(2-hydroxyphenyl)acetamide (NA-2) and Temozolomide synergistically induce apoptosis in human glioblastoma cell line U87	Hanif et al.	2014	10.1186/s12935-014-0133-5	TRUE
65	Polyphyllin VII Promotes Apoptosis and Autophagic Cell Death via ROS-Inhibited AKT Activity and Sensitizes Glioma Cells to Temozolomide	Pang et al.	2019	10.1155/2019/1805635	TRUE
66	Quercetin sensitizes human glioblastoma cells to temozolomide in vitro via inhibition of Hsp27	Sang et al.	2014	10.1038/aps.2014.22	TRUE
67	Radiobiological evaluation and correlation with the local effect model (LEM) of carbon ion radiation therapy and temozolomide in glioblastoma cell lines	Combs et al.	2008	10.1080/09553000802641151	TRUE
68	Receptor-mediated PLGA nanoparticles for glioblastoma multiforme treatment	Ramalho et al.	2018	10.1016/j.ijpharm.2018.04.062	TRUE
69	Regulation of Integrated Stress Response Sensitizes U87MG Glioblastoma Cells to Temozolomide Through the Mitochondrial Apoptosis Pathway	He et al.	2018	10.1002/ar.23839	TRUE
70	Riluzole enhances the antitumor effects of temozolomide via suppression of MGMT expression in glioblastoma	Yamada et al.	2020	10.3171/2019.12.JNS192682	TRUE

71	Salvianolic acid B renders glioma cells more sensitive to radiation via Fis-1-mediated mitochondrial dysfunction	Chen et al.	2018	10.1016/j.biopha.2018.08.113	TRUE
72	Sequence-dependent synergistic inhibition of human glioma cell lines by combined temozolomide and miR-21 inhibitor gene therapy	Qian et al.	2012	10.1021/mp3002039	TRUE
73	Sequential treatment of phenethyl isothiocyanate increases sensitivity of temozolomide resistant glioblastoma cells by decreasing expression of mgmt via nf-kappab pathway	Guo et al.	2019		TRUE
74	Silencing SATB1 overcomes temozolomide resistance by downregulating MGMT expression and upregulating SLC22A18 expression in human glioblastoma cells	Yang et al.	2018	10.1038/s41417-018-0040-3	TRUE
75	Sirtuin 1 knockdown inhibits glioma cell proliferation and potentiates temozolomide toxicity via facilitation of reactive oxygen species generation	Chen et al.	2019	10.3892/ol.2019.10235	TRUE
76	STAT3 Inhibition Overcomes Temozolomide Resistance in Glioblastoma by Downregulating MGMT Expression	Kohsaka et al.	2012	10.1158/1535-7163.MCT-11-0801	TRUE
77	Study on therapeutic action and mechanism of TMZ Combined with RITA against glioblastoma	Wu et al.	2018	10.1159/000495923	TRUE
78	Suppression of the Eag1 potassium channel sensitizes glioblastoma cells to injury caused by temozolomide	Sales et al.	2016	10.3892/ol.2016.4992	TRUE

79	Synergistic inhibition of human glioma cell line by temozolomide and PAMAM-mediated miR-21i	Qian et al.	2012	10.1002/app.37823	TRUE
80	Synergistic suppression of noscapine and conventional chemotherapeutics on human glioblastoma cell growth	QI et al.	2013	10.1038/aps.2013.40	TRUE
81	Targeted Brain Tumor Therapy by Inhibiting the MDM2 Oncogene: In Vitro and In Vivo Antitumor Activity and Mechanism of Action	Punganuru et al.	2020	10.3390/cells9071592	TRUE
82	TAZ promotes temozolomide resistance by upregulating MCL-1 in human glioma cells	Tian et al.	2015	10.1016/j.bbrc.2015.05.115	TRUE
83	Temozolomide Cocrystals Exhibit Drug Sensitivity in Glioblastoma Cells	Kusuma et al.	2014	10.1007/s40010-014-0142-8	TRUE
84	Temozolomide induces autophagy via ATM-AMPK-ULK1 pathways in glioma	Zou et al.	2014	10.3892/mmr.2014.2151	TRUE
85	The DNA repair protein ALKBH2 mediates temozolomide resistance in human glioblastoma cells	Johannessen et al.	2012	10.1093/neuonc/nos301	TRUE
86	The Effect of Ascorbic Acid over the Etoposide- and Temozolomide-Mediated Cytotoxicity in Glioblastoma Cell Culture: A Molecular Study	Gokturk et al.	2018	10.5137/1019-5149.JTN.19111-16.1	TRUE
87	The effect of polysaccharides from Cibotium barometz on enhancing temozolomide-induced glutathione exhausted in human glioblastoma U87 cells as revealed by H-1 NMR metabolomics analysis	Shi et al.	2020	10.1016/j.ijbiomac.2020.03.243	TRUE

88	The effect of silibinin in enhancing toxicity of temozolomide and etoposide in p53 and PTEN-mutated resistant glioma cell lines	Elhag et al.	2015		TRUE
89	The Effect of Temozolomide/Poly(lactide-co-glycolide) (PLGA)/Nano-Hydroxyapatite Microspheres on Glioma U87 Cells Behavior	Zhang et al.	2012	10.3390/ijms13011109	TRUE
90	The HIV-derived protein Vpr52-96 has anti-glioma activity in vitro and in vivo	Kübler et al.	2016	10.1158/1538-7445.AM2015-4458	TRUE
91	The mTOR inhibitor RAD001 potentiates autophagic cell death induced by temozolomide in a glioblastoma cell line	Josset et al.	2013		TRUE
92	The Pan-Bcl-2 Inhibitor (-)-Gossypol Triggers Autophagic Cell Death in Malignant Glioma	Voss et al.	2010	10.1158/1541-7786.MCR-09-0562	TRUE
93	The Synergistic Effect of Combination Progesterone and Temozolomide on Human Glioblastoma Cells	Atif et al.	2015	10.1371/journal.pone.0131441	TRUE
94	The synergistic effect of combination temozolomide and chloroquine treatment is dependent on autophagy formation and p53 status in glioma cells	Lee et al.	2015	10.1016/j.canlet.2015.02.012	TRUE
95	Tim-3 expression in glioma cells is associated with drug resistance	Zhang et al.	2019	10.4103/jcrt.JCRT_630_18	TRUE
96	Tramadol attenuates the sensitivity of glioblastoma to temozolomide through the suppression of Cx43-mediated gap junction intercellular communication	Wang et al.	2018	10.3892/ijo.2017.4188	TRUE

97	Transcriptional targeting of adenovirally delivered tumor necrosis factor alpha by temozolomide in experimental glioblastoma	Yamini et al.	2004	10.1158/0008-5472.CAN-04-2117	TRUE
98	Verapamil potentiates anti-glioblastoma efficacy of temozolomide by modulating apoptotic signaling	Hanif et al.	2018	10.1016/j.tiv.2018.07.001	TRUE
99	Verbascoside inhibits glioblastoma cell proliferation migration and invasion while promoting apoptosis through upregulation of protein tyrosine phosphatase SHP-1 and inhibition of STAT3 phosphorylation	Jia et al.	2018	10.1159/000491067	TRUE
100	Zinc enhances TMZ cytotoxicity in glioblastoma multiforme model system	Toren et al.	2016	10.1093/neuonc/not176	TRUE
101	β-Elemene inhibits proliferation through crosstalk between glia maturation factor β and extracellular signal-regulated kinase 1/2 and impairs drug resistance to temozolomide in glioblastoma cells	Zhu et al.	2014	10.3892/mmr.2014.2273	TRUE
102	Anti-epidermal growth factor receptor siRNA carried by chitosan-transacylated lipid nanocapsules increases sensitivity of glioblastoma cells to temozolomide	Messaoudi et al.	2014	10.2147/IJN.S59134	FALSE
103	Autophagic flux response and glioblastoma sensitivity to radiation	Mitrakas et al.	2018	10.20892/j.issn.2095-3941.2017.0173	FALSE
104	Autophagy mediates glucose starvation-induced glioblastoma cell quiescence and chemoresistance through coordinating cell metabolism cell cycle and survival	Wang et al.	2018	10.1038/s41419-017-0242-x	FALSE

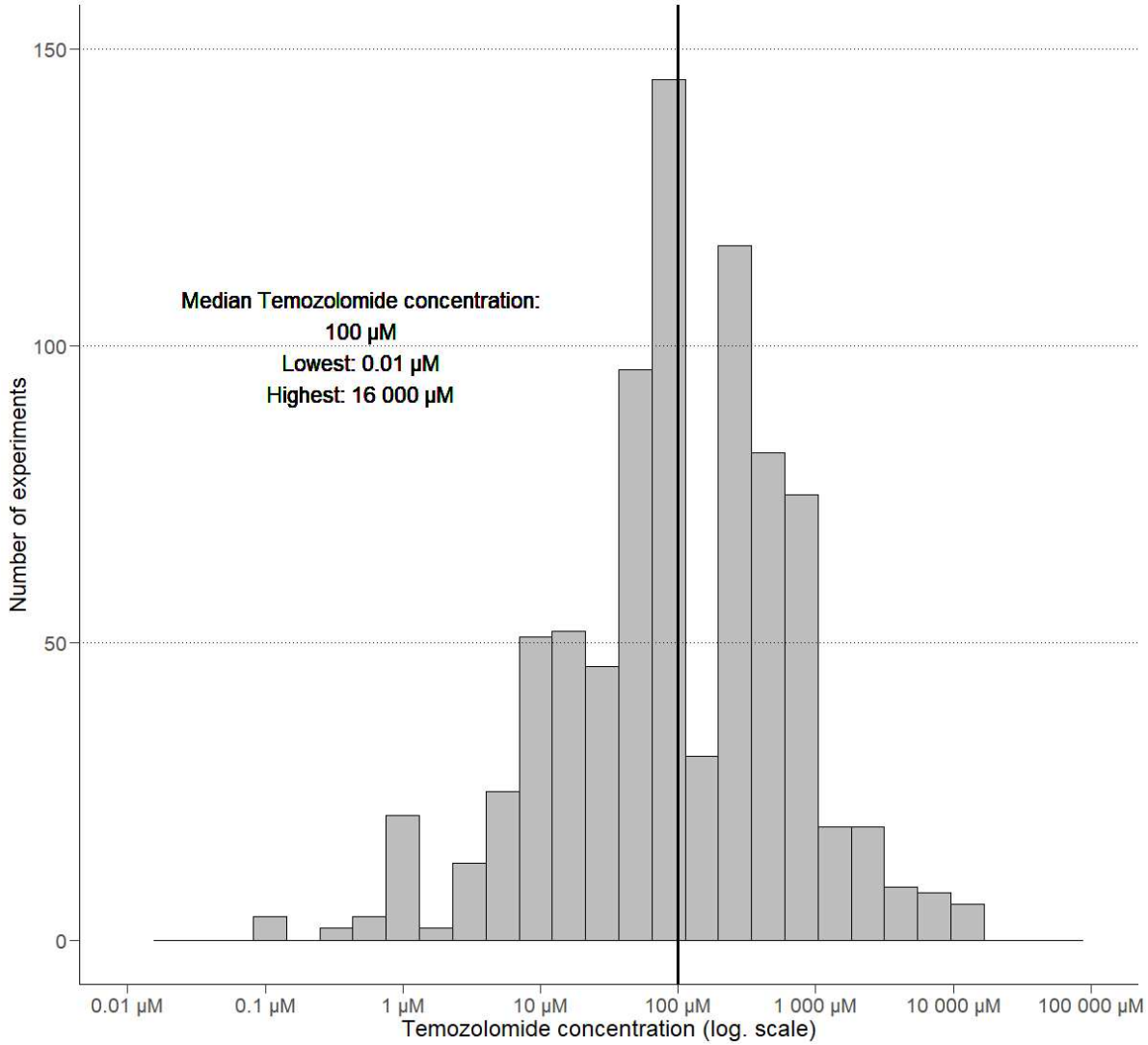
105	BET inhibitor I-BET151 sensitizes GBM cells to temozolomide via PUMA induction	Yao et al.	2019	10.1038/s41417-018-0068-4	FALSE
106	Combined effect of 2-5A-linked antisense against telomerase RNA and conventional therapies on human malignant glioma cells in vitro and in vivo	Iwado et al.	2007	10.3892/ijo.31.5.1087	FALSE
107	Decreasing GSH and increasing ROS in chemosensitivity gliomas with IDH1 mutation	Shi et al.	2014	10.1007/s13277-014-2644-z	FALSE
108	Downregulation of Id2 increases chemosensitivity of glioma	Zhao et al.	2015	10.1007/s13277-015-3055-5	FALSE
109	Downregulation of miR-155 inhibits proliferation and enhances chemosensitivity to Temozolomide in glioma cells	Meng et al.	2017		FALSE
110	Downregulation of miR-196b Promotes Glioma Cell Sensitivity to Temozolomide Chemotherapy and Radiotherapy	Ma et al.	2018		FALSE
111	Effect of temozolomide on the viability of musculoskeletal sarcoma cells	Kusabe et al.	2015	10.3892/ol.2015.3506	FALSE
112	Effects of galbanic acid on proliferation migration and apoptosis of glioblastoma cells through PI3K/Akt/mTOR signaling pathway	Shahcheraghi et al.	2020	10.2174/1874467213666200512075507	FALSE
113	Encapsulation of Temozolomide in a Calixarene Nanocapsule Improves Its Stability and Enhances Its Therapeutic Efficacy against Glioblastoma	Renziehausen et al.	2019	10.1158/1535-7163.MCT-18-1250	FALSE
114	FTO inhibition enhances the anti-tumor effect of temozolomide by targeting MYC-miR-155/23a cluster-MXI1 feedback circuit in glioma	Xiao et al.	2020	10.1158/0008-5472.CAN-20-0132	FALSE

115	Green tea epigallocatechin gallate enhances therapeutic efficacy of temozolomide in orthotopic mouse glioblastoma models	Chen et al.	2011	10.1016/j.canlet.2010.11.008	FALSE
116	Growth-inhibitory and chemosensitizing effects of microRNA-31 in human glioblastoma multiforme cells	Zhou et al.	2015	10.3892/ijmm.2015.2312	FALSE
117	Growth-inhibitory and chemosensitizing effects of the glutathione-S- transferase-pi-activated nitric oxide donor PABA/NO in malignant gliomas	Kogias et al.	2012	10.1002/ijc.26106	FALSE
118	Identification of Key Candidate Proteins and Pathways Associated with Temozolomide Resistance in Glioblastoma Based on Subcellular Proteomics and Bioinformatical Analysis	Yi et al.	2018	10.1155/2018/5238760	FALSE
119	Inhibition of EZH2 reverses chemotherapeutic drug TMZ chemosensitivity in glioblastoma	Fan et al.	2014		FALSE
120	Inhibition of JNK Potentiates Temozolomide-induced Cytotoxicity in U87MG Glioblastoma Cells via Suppression of Akt Phosphorylation	Vo et al.	2014		FALSE
121	Long noncoding RNA RP11-838N2.4 enhances the cytotoxic effects of temozolomide by inhibiting the functions of miR-10a in glioblastoma cell lines	Liu et al.	2016	10.18632/oncotarget.9699	FALSE
122	LY294002 enhances cytotoxicity of temozolomide in glioma by down-regulation of the PI3K/Akt pathway	Chen et al.	2012	10.3892/mmr.2011.674	FALSE
123	Mechanisms operative in the antitumor activity of temozolomide in glioblastoma multiforme	Fischer et al.	2007	10.1097/PPO.0b013e318157053f	FALSE

124	MiR-21 protected human glioblastoma U87MG cells from chemotherapeutic drug temozolomide induced apoptosis by decreasing Bax/Bcl-2 ratio and caspase-3 activity	Shi et al.	2010	10.1016/j.brainres.2010.07.009	FALSE
125	Mutant TP53 enhances the resistance of glioblastoma cells to temozolomide by up-regulating O-6-methylguanine DNA-methyltransferase	Wang et al.	2012	10.1007/s10072-012-1257-9	FALSE
126	Next Generation Sequencing-Based Transcriptome Predicts Bevacizumab Efficacy in Combination with Temozolomide in Glioblastoma	Adilijiang et al.	2019	10.3390/molecules24173046	FALSE
127	Olanzapine inhibits proliferation migration and anchorage-independent growth in human glioblastoma cell lines and enhances temozolomide's antiproliferative effect	Karpel-Massler et al.	2014	10.1007/s11060-014-1688-7	FALSE
128	Ovatodiolide inhibits the oncogenicity and cancer stem cell-like phenotype of glioblastoma cells as well as potentiate the anticancer effect of temozolomide	Su et al.	2019	10.1016/j.phymed.2019.152840	FALSE
129	Overexpression of iASPP-SV in glioma is associated with poor prognosis by promoting cell viability and antagonizing apoptosis	Liu et al.	2015	10.1007/s13277-015-4503-y	FALSE
130	Polymer - Temozolomide Conjugates as Therapeutics for Treating Glioblastoma	Ward et al.	2018	10.1021/acs.molpharmaceut.8b00766	FALSE
131	Potentiation of antiglioma effect with combined temozolomide and interferon-beta	Park et al.	2006		FALSE
132	Silence of bFGF enhances chemosensitivity of glioma cells to temozolomide through the MAPK signal pathway	Wang et al.	2016	10.1093/abbs/gmw035	FALSE

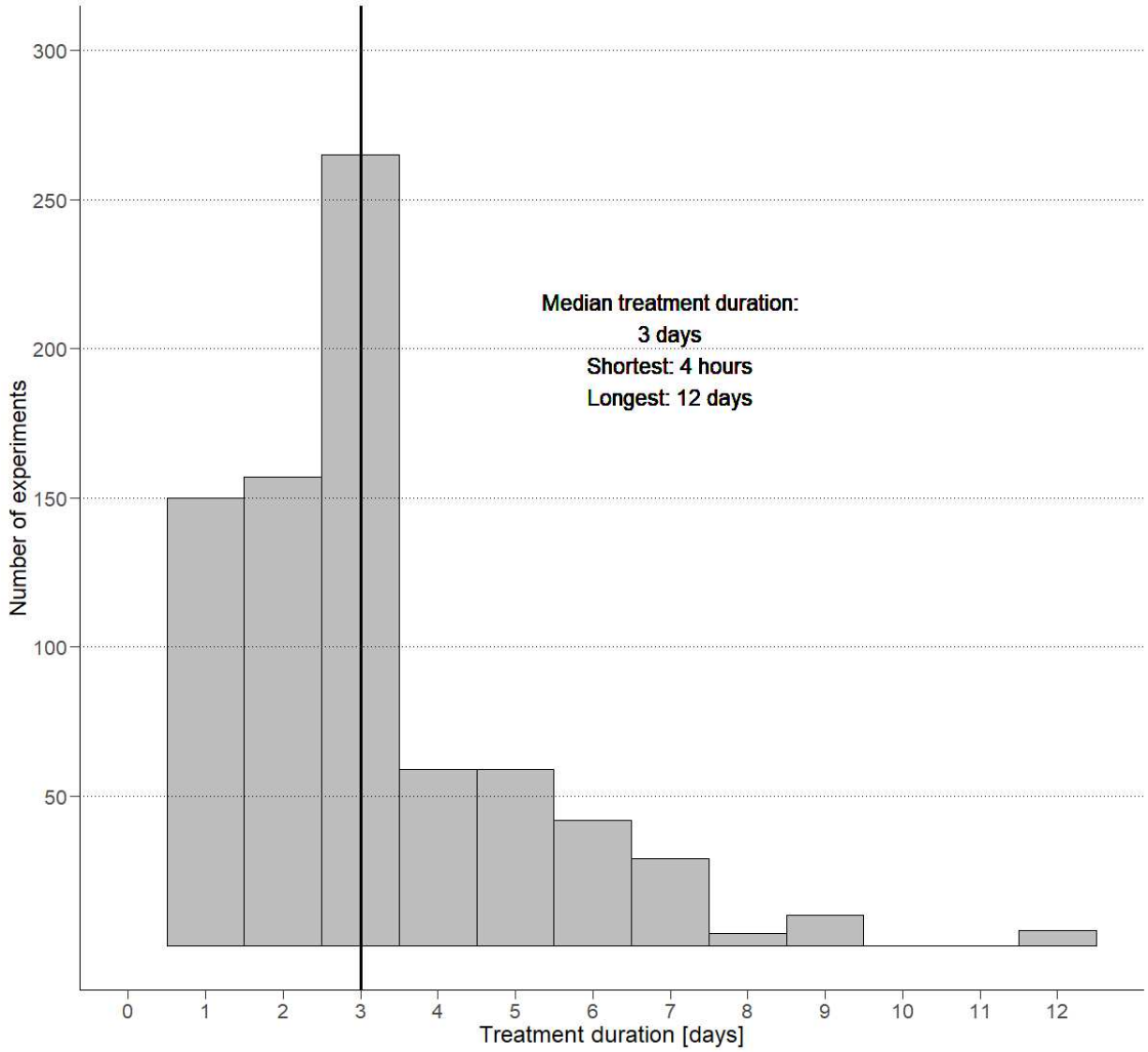
133	Synergistic combination of chemo-phototherapy based on temozolomide/ICG-loaded iron oxide nanoparticles for brain cancer treatment	Kwon et al.	2019	10.3892/or.2019.7289	FALSE
134	Temozolomide Gemcitabine and Decitabine Hybrid Nanoconjugates: From Design to Proof-of-Concept (PoC) of Synergies toward the Understanding of Drug Impact on Human Glioblastoma Cells	Sahli et al.	2020	10.1021/acs.jmedchem.0c00694	FALSE
135	The synergic antitumor effects of paclitaxel and temozolomide co-loaded in mPEG-PLGA nanoparticles on glioblastoma cells	Xu et al.	2016	10.18632/oncotarget.7896	FALSE
136	Vincristine and temozolomide combined chemotherapy for the treatment of glioma: a comparison of solid lipid nanoparticles and nanostructured lipid carriers for dual drugs delivery	Wu et al.	2015	10.3109/10717544.2015.1058434	FALSE
137	YKL-40 downregulation is a key factor to overcome temozolomide resistance in a glioblastoma cell line	Akiyama et al.	2014	10.3892/or.2014.3195	FALSE

7.1.7 S7: Reported temozolomide concentrations in the meta-analysis



The 137 articles included cell viability data for 828 temozolomide concentrations (98 unique). The drug's concentration is presented on a logarithmic scale.

7.1.8 S8: Reported treatment durations in the included studies into meta-analysis



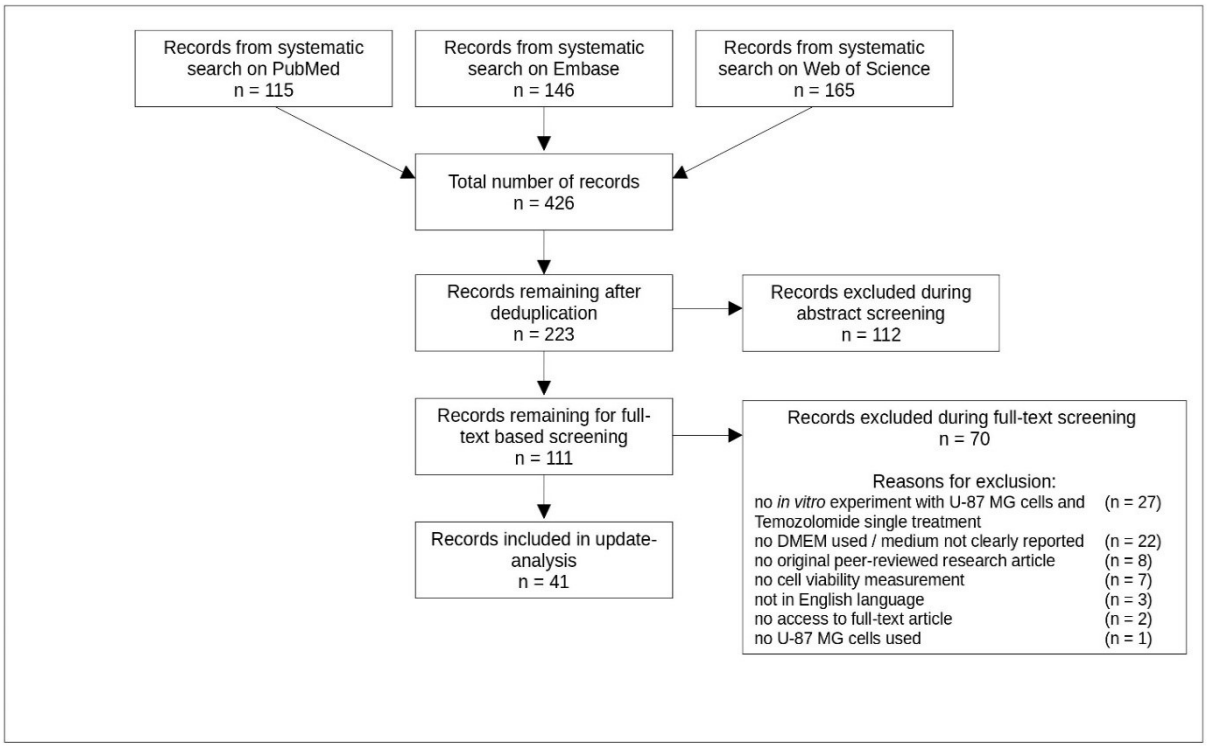
The 137 articles included cell viability data for 786 durations of temozolomide exposure (20 unique). The lower number in comparison to used temozolomide concentrations is due to non-reporting. Treatment duration was defined as the duration of exposure of Uppsala-87 Malignant Glioma cells to temozolomide.

7.1.9 S9: Risk of bias factors prevalence

Potential risk of bias	Articles	
Missing sample size calculation	137	100.0%
No random group allocation	137	100.0%
No blinded outcome assessment	137	100.0%
No open-access study protocol available	137	100.0%
Unclear way of calculation for cell viability average and error values	92	67.2%
Unclear number of independent experiments and replications per experiment	47	34.3%
Data were not presented for every experiment	12	8.8%
Missing data (for particular drug concentrations and/or treatment durations) within an experiment	1	0.7%

The column "articles" shows the absolute and relative prevalence of articles having a particular risk of bias factor in comparison to all 137 included articles

7.1.10 S10: Literature flow chart of systematic search update



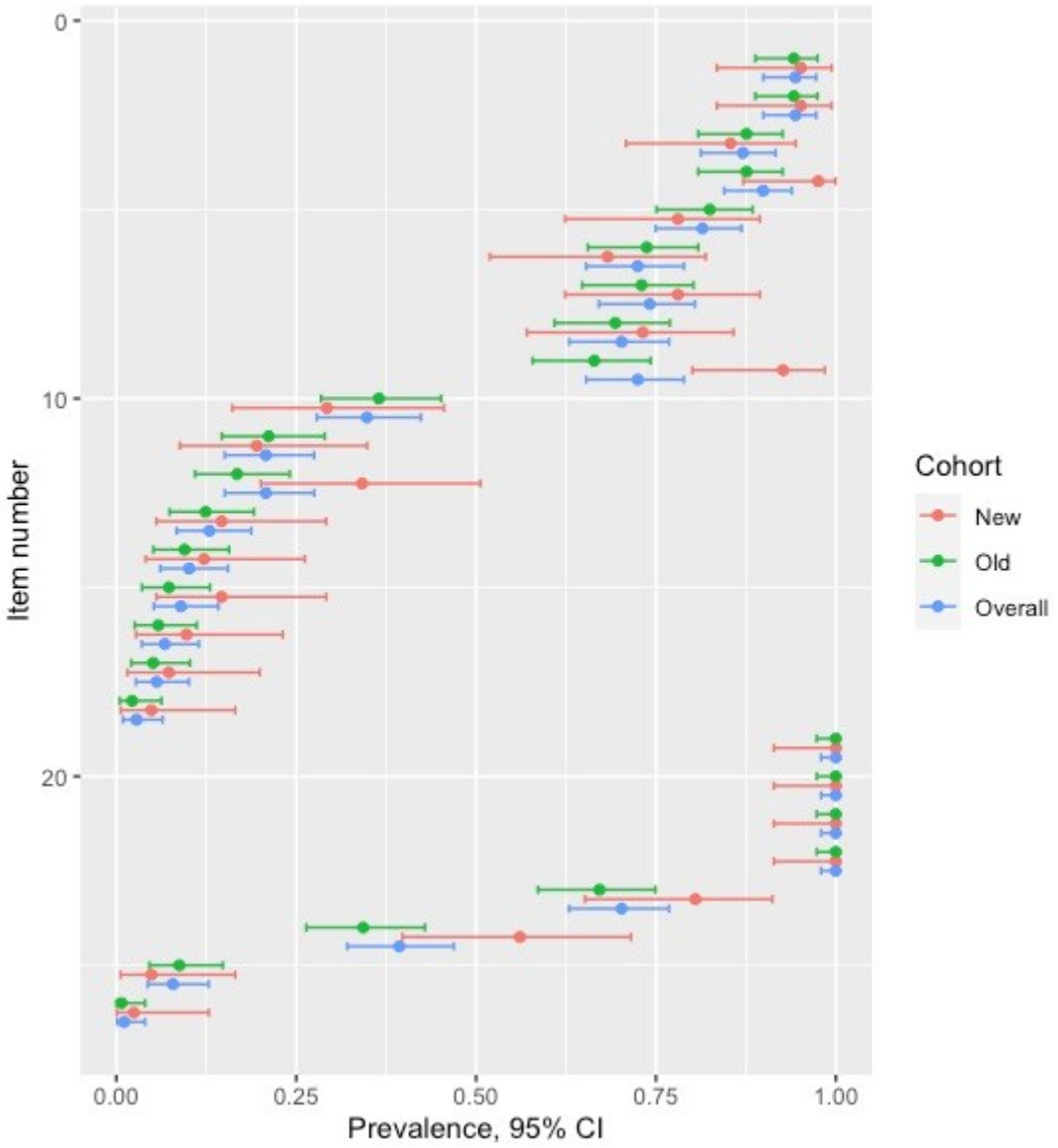
Presentation based on the PRISMA statement (2020). Updated systematic search was conducted on 10th of March 2022. One reason for exclusion per excluded article. DMEM = Dulbecco's modified Eagle's medium; U-87 MG = Uppsala-87 Malignant Glioma.

7.1.11 S11: Updated assessment of reporting and risks of bias

	Study parameters‘ reporting quality			
No.	**Parameter**	**Reporting 09/2020 – 03/2022**	**Reporting until 08/2020**	**Change**
1	Source of U-87 MG	95.1%	94.2%	0.9%
2	Concentration of TMZ	95.1%	94.2%	0.9%
3	Errortype	85.4%	87.6%	-2.2%
4	Treatment duration	97.6%	87.6%	10.0%
5	Number of experiments	78.0%	82.5%	-4.5%
6	Source of TMZ	68.3%	73.7%	-5.4%
7	Antibiotics	78.0%	73.0%	5.0%
8	Source of FBS	73.2%	69.3%	3.9%
9	Conflicts of interest	92.7%	66.4%	26.3%
10	Type of control	29.3%	36.5%	-7.2%
11	Glucose level	19.5%	21.2%	-1.7%
12	U-87 MG concentration	34.1%	16.8%	17.3%
13	Cell passaging criteria	14.6%	12.4%	2.2%
14	U-87 MG authentication	12.2%	9.5%	2.7%
15	U-87 MG age	14.6%	7.3%	7.3%
16	Mycoplasma exclusion	9.8%	5.8%	4.0%
17	Volume of added TMZ	7.3%	5.1%	2.2%
18	Volume of added control	4.9%	2.2%	2.7%
	Prevalence of risks of bias			
	Risk of bias	**Prevalence 09/2020 – 03/2022**	**Prevalence until 08/2020**	**Change**
19	Missing sample size calculation	100.0%	100.0%	0.0%
20	No random group allocation	100.0%	100.0%	0.0%
21	No blinded outcome assessment	100.0%	100.0%	0.0%
22	No open-access study protocol available	100.0%	100.0%	0.0%
23	Unclear way of calculation for cell viability average and error values	80.5%	67.2%	13.3%
24	Unclear number of independent experiments and replications per experiment	56.1%	34.3%	21.8%
25	Data were not presented for every experiment	4.9%	8.8%	-3.9%
26	Missing data (for particular drug concentrations and/or treatment durations) within an experiment	2.4%	0.7%	1.7%

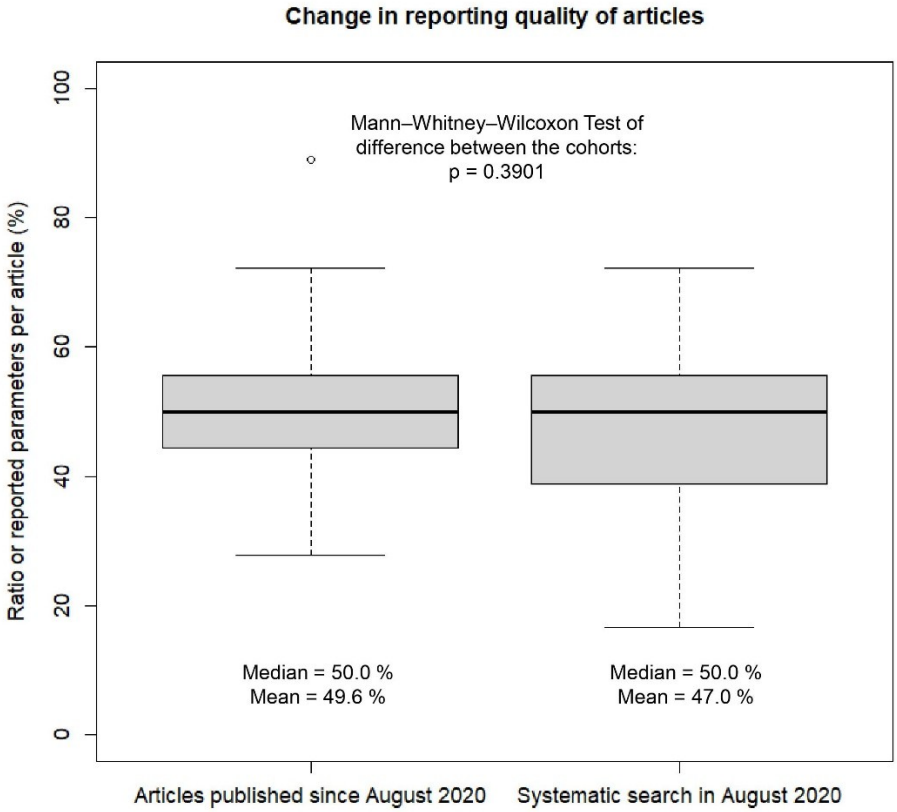
Comparison of study parameter reporting quality (share of articles that reported the parameter) and prevalence of risks of bias until the time of the systematic literature search in 03/2020 (137 articles) and thereafter until 03/2022 (41 articles). Change (%) was calculated setting reporting/prevalence of risks of bias until 08/2020 as reference point (100 %). FBS = fetal bovine serum, No. = number, TMZ = Temozolomide, U-87 MG = Uppsala 87 Malignant Glioma.

7.1.12 S12: Reliability of the updated assessment of reporting and risks of bias



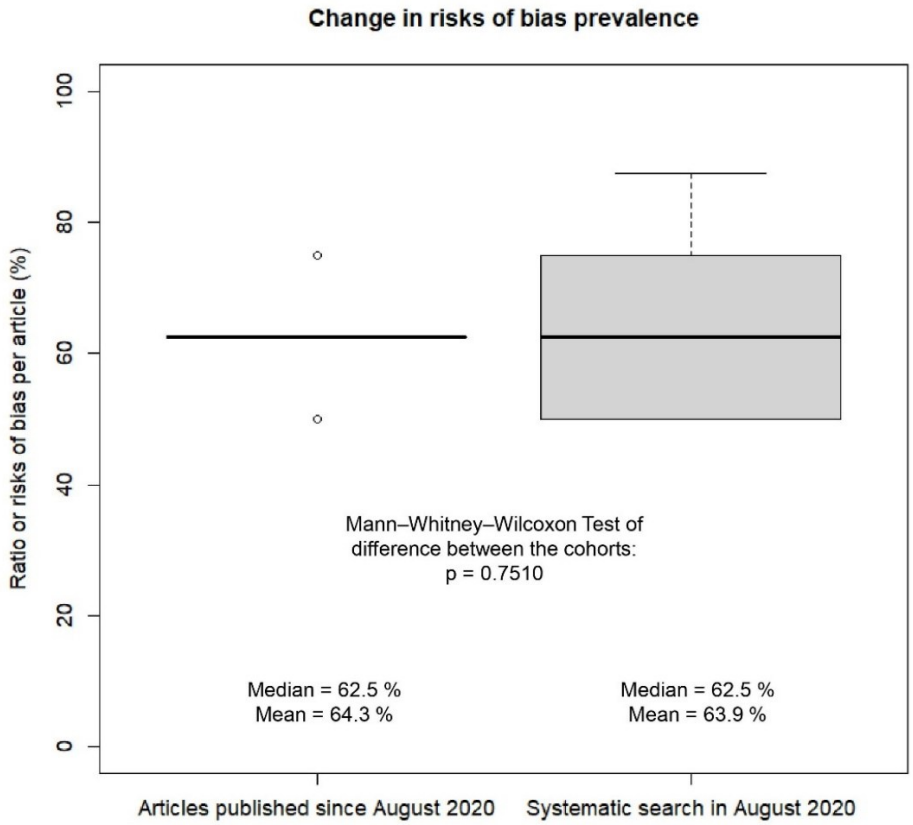
Confidence intervals (95 %) for the prevalence of reported study parameters and risks of bias. Y-axis indicates the item number according to S11. Red data indicate prevalence in the search for additional articles in the timespan of September 2020 until March 2022, green data indicate prevalence in the systematic search conducted in August 2020, and the blue data indicate the overall prevalence including old and new search results.

7.1.13 S13: Update on overall reporting quality of articles



Comparison of articles reporting quality at the time points of initial systematic search in August 2020 and in March 2022. Article reporting quality was calculated as the share of reported parameters including the 18 study parameters presented in Supplement S10. Median and mean reporting qualities for both cohorts are presented underneath the boxplots. The boxplots indicate median, 25th (Q1) and 75th (Q3) percentiles (interquartile range (IQR)), Q1 – 1.5 x IQR, Q3 + 1.5 x IQR, and outliers of reporting quality. Mann-Whitney-Wilcoxon test was used for comparison of reporting across both cohorts with the null hypothesis of no change in reporting.

7.1.14 S14: Update on overall risks of bias prevalence



Comparison of the prevalence of risks of bias at the time points of initial systematic search in August 2020 and in March 2022. The prevalence of risks of bias was calculated as the share of present risks of bias including the eight potential risk of bias parameters presented in Supplement S10. Median and mean risks of bias prevalence for both cohorts are presented underneath the boxplots. The boxplots indicate median, 25th (Q1) and 75th (Q3) percentiles (interquartile range (IQR)), Q1 – 1.5 x IQR, Q3 + 1.5 x IQR, and outliers of reporting quality. Mann-Whitney-Wilcoxon test was used for comparison of risks of bias prevalence across both cohorts with the null hypothesis of no change in prevalence.

7.1.15 S15: Moderators of within-articles-variance of true effects

Moderator	Type	Number of effects	Number of articles	*p* value	Marginal R^2	Within-articles-variance		
						tau^2	I^2	Explained
Without moderators		644	101			3.6%	56.6%	n. a.
U-87 MG source	cat.	644	101	.075	n. s.	3.6%		
U-87 MG authentication	cat.	644	101	.476	n. s.	3.6%		
U87-MG age (Cell passages)	cont.	138	11	.238	n. s.	4.0%		
Cell concentration	cont.	113	20	.323	n. s.	4.8%		
Confluence level at cell passaging	cont.	57	11	.319	n. s.	4.2%		
Glucose level of culture medium	cat.	644	101	.016	7.0%	3.6%[a]	59.4%	0.0%
Mycoplasma exclusion	cat.	644	101	.491	n. s.	3.6%		
Supplemented antibiotics	cat.	644	101	.094	n. s.	3.6%		
FBS source	cat.	644	101	.067	n. s.	3.6%		
Type of control	cat.	644	101	.370	n. s.	3.6%		
Articles reporting quality	int.	644	101	.031	3.3%	3.6%[b]	57.8%	0.0%
TMZ conc.	cont.	644	101	< .001	38.6 %	1.9%[c]	35.0%	48.3%
Treatment duration	cont.	644	101	< .001	6.0%	3.4%[d]	53.8%	4.8%

Random-effects three-level meta-regressions with the raw data the effects were calculated with as first level, the reported effects as second level and the articles the effects were reported in as third level. Marginal R^2 indicates the regression model fit (262); tau^2: estimator of the variance of true effects; tau = square root of tau^2; I^2: proportion of within-articles-variance of the total observed variance including sampling error. tau^2 estimator: restricted-maximum likelihood. The column "explained" indicates the reduction of tau^2 after including the particular moderator compared to tau^2 without moderators (only applicable if the number of included effects and articles is identical). Types of moderators: cat. = categorical; cont. = continuous, int. = interval. For some continuous moderators, the number of effects and articles included in the regression is reduced due to non-reporting which leads to a limited comparability of tau^2 between parameters with different numbers of belonging articles and effects. "Not reported" was included as a category for categorical moderators. The *p* value is for the test of the moderator. R^2, I^2 and the explained heterogeneity were only calculated for moderators that prove significance in the test of the moderator (alpha = .05). TMZ conc. = Temozolomide concentration. **a**: 95%-confidence-interval (CI): tau^2: [3.2%,4.1%]; **b**: 95%-CI of tau^2: [3.2%,4.1%]; **c**: 95%-CI of tau^2: [1.6%,2.2%]; **d**: 95%-CI of tau^2: [3.0%,3.9%].

7.1.16 S16: Correlation between study parameters and the TMZ sensitivity

Moderator	Effects	Articles	Cell viability reduction	SD	CI LL	CI UL	*p* value
Glucose level							
High glucose (4500 mg/dl)	168	21	23.1%	3.6%	15.8%	30.5%	< 0.001
Low glucose (1000 mg/dl)	37	2	20.6%	5.8%	-25.2%	66.3%	.726*
No glucose	1	1	29.0%	3.5%	21.6%	36.4%	.113*
Not reported	438	77	37.1%	4.2%	28.6%	45.6%	.002*
Articles reporting quality							
Intercept	644	101	61.3%	12.0%	37.0%	85.7%	< .001
Change per unit increase in the number of reported parameters			-3.0%	1.3%	-5.5%	-0.4%	.026

Univariable Random-effects three-level meta-regressions with the raw data the effects were calculated with as first level, the reported effects as second level and the articles the effects were reported in as third level. Effects were estimated using robust-variance-estimation. Cell viability reduction is presented in comparison to the corresponding untreated control. A linear regression model was applied for the articles reporting quality correlation analysis. The reduced number of included articles and effects for glucose level analysis was a result of non-reporting. TMZ = Temozolomide; SD = standard deviation; CI = confidence interval (with significance level of alpha = .05); LL = lower limit; UL = upper limit. [a]: The *p* value indicates whether there was a significant difference for the parameter phenotypes effect estimate in comparison to the effect estimate in the high glucose group.

7.2 Supplementary material to publication 2

The following supplementary material has been published alongside the abovementioned publication (2). The numeration of supplements has been maintained according to the original numeration used in the publication.

7.2.1 S1: Literature screening criteria

Inclusion criteria	**Exclusion criteria**
SAH perforation model	Other SAH models (e.g., injection model)
Wild-type mice as experimental animals	Other animals (e.g., rats) Genetically modified mice
Reporting of at least one outcome: animal mortality, SAH severity grade, large artery vasospasm	None of the outcomes reported
Original peer-reviewed research articles	Other publication types (e.g., conference abstracts, book chapters)
English language	Other languages than English

All inclusion criteria had to be fulfilled for an article to be included in the review. SAH = Subarachnoid haemorrhage.

7.2.2 S2: Extracted parameters

Category	**Parameters**
Mice as experimental animals	Number Strain Sex Age Weight
Animal housing conditions	Single cage vs. group cage Temperature Humidity Twelve-hour light-dark-cycle Free access to food and water
Anaesthesia	Inhalative or injective anaesthesia Inhalative/Injective anaesthetic drugs Inhalative anaesthesia O2/N2-ratio and flow Intubation
Surgery	Perforation entry point Location of perforation Duration of SAH perforation surgery Monitoring of ICP Postsurgical pain/stress-management
Filament for perforation	Diameter Length Material Tip (sharpened vs. blunted) Texture: monofilament vs. suture
Outcome	Mortality rate, SAH severity grade, large artery vasospasm Timespan after SAH perforation Vehicle controls

ICP = Intracranial pressure, SAH = Subarachnoid haemorrhage.

7.2.3 S3: Table of all included articles

Title	Author and year	DOI	Mortality	SAH grade	Vasospasm
A segmentation-based volumetric approach to localize and quantify cerebral vasospasm based on tomographic imaging data	Neulen (2017)	10.1371/journal.pone.0172010	30.8 %	NA	NA
An apoE-derived mimic peptide, COG1410, alleviates early brain injury via reducing apoptosis and neuroinflammation in a mouse model of subarachnoid hemorrhage	Wu (2016)	10.1016/j.neuret.2016.05.058	28.6 %	13.3 (SD =2.8)	NA
Analgesic treatment limits surrogate parameters for early stress and pain response after experimental subarachnoid hemorrhage	Staib-Lasarzik (2019)	10.1186/s12868-019-0531-7	37.5 %	NA	NA
Anesthetic and subanesthetic doses of isoflurane conditioning provides strong protection against delayed cerebral ischemia in a mouse model of subarachnoid hemorrhage	Athiraman (2021)	10.1016/j.brainres.2020.147169	10.5 %	NA	69.6 % (SD = 22.0 %)
Anti-Vascular Endothelial Growth Factor Treatment Suppresses Early Brain Injury After Subarachnoid Hemorrhage in Mice	Liu (2016)	10.1007/s12035-015-9386-9	23.1 %	11.0 (1.3)	NA
Apolipoprotein E Deficiency Aggravates Neuronal Injury by Enhancing Neuroinflammation via the JNK/c-Jun Pathway in the Early Phase of Experimental Subarachnoid Hemorrhage in Mice	Wu (2019)	10.1155/2019/3832648	29.4 %	11.5 (2.5)	NA
Biglycan regulates neuroinflammation by promoting M1 microglial activation in early brain injury after experimental subarachnoid hemorrhage	Xie (2020)	10.1111/jnc.14926	21.4 %	12.9 (1.5)	NA
Calcium sensing receptor contribute to early brain injury through the CaMKII/NLRP3 pathway after subarachnoid hemorrhage in mice	Wang (2020)	10.1016/j.bbrc.2020.07.081	17.2 %	10.3 (2.2)	NA

Title	Author and year	DOI	Mortality	SAH grade	Vasospasm
Capillary flow disturbances after experimental subarachnoid hemorrhage: A contributor to delayed cerebral ischemia?	Anzabi (2019)	10.1111/micc.12516	61.0 %	NA	NA
Deficiency of Tenascin-C Alleviates Neuronal Apoptosis and Neuroinflammation After Experimental Subarachnoid Hemorrhage in Mice	Liu (2018)	10.1007/s12035-018-1006-z	26.7 %	10.0 (4.7)	NA
DHEA Attenuates Microglial Activation via Induction of JMJD3 in Experimental Subarachnoid Haemorrhage	Tao (2019)	10.1186/s12974-019-1641-y	18.4 %	NA	NA
Effect of ADAMTS-13 on cerebrovascular microthrombosis and neuronal injury after experimental subarachnoid hemorrhage	Muroi (2014)	10.1111/jth.12511	22.4 %	NA	NA
Effect of decompressive craniectomy on outcome following subarachnoid hemorrhage in mice	Buehler (2015)	10.1161/STROKEAHA.114.007703	10.0 %	NA	NA
Effects of Toll-Like Receptor 4 Antagonists Against Cerebral Vasospasm After Experimental Subarachnoid Hemorrhage in Mice	Kawakita (2017)	10.1007/s12035-016-0178-7	13.3 %	8.6 (0.9)	49.9 % (13.6 %)
Endothelial nitric oxide synthase mediates endogenous protection against subarachnoid hemorrhage-induced cerebral vasospasm	Vellimana (2011)	10.1161/STR.0b013e3182074d88	5.8 %	NA	73.4 % (24.0 %)
Endovascular Perforation Murine Model of Subarachnoid Hemorrhage	Du (2016)	10.1007/978-3-319-18497-5_14	33.3 %	NA	NA
Evaluation of a filament perforation model for mouse subarachnoid hemorrhage using 7.0 Tesla MRI	Muroi (2016)	10.1016/j.jocn.2015.10.045	90.9 %	NA	NA
Filament perforation model for mouse subarachnoid hemorrhage: surgical-technical considerations	Muroi (2014)	10.3109/02688697.2014.918579	21.1 %	NA	NA
Inhibition of AMPA (alpha-Amino-3-Hydroxy-5-Methyl-4-Isoxazole Propionate) Receptor Reduces Acute Blood-Brain Barrier Disruption After	Kawakita (2021)	10.1007/s12975-021-00934-0	18.5 %	8.9 (2.3)	NA

Title	Author and year	DOI	Mortality	SAH grade	Vasospasm
Subarachnoid Hemorrhage in Mice					
Integrated analysis of gait parameters and gene expression profiles in a murine model of subarachnoid hemorrhage	Zheng (2021)	10.1111/gbb.12728	20.0 %	NA	NA
Irisin Contributes to Neuroprotection by Promoting Mitochondrial Biogenesis After Experimental Subarachnoid Hemorrhage	Tu (2021)	10.3389/fnagi.2021.640215	18.1 %	14.7 (1.7)	NA
Long-term impairment of neurovascular coupling following experimental subarachnoid hemorrhage	Balbi (2020)	10.1177/0271678X19863021	20.0 %	NA	NA
Melatonin Attenuates Early Brain Injury via the Melatonin Receptor/Sirt1/NF-κB Signaling Pathway Following Subarachnoid Hemorrhage in Mice	Zhao (2017)	10.1007/s12035-016-9776-7	20.0 %	12.1 (2)	NA
Melatonin Attenuates White Matter Injury via Reducing Oligodendrocyte Apoptosis After Subarachnoid Hemorrhage in Mice	Liu (2020)	10.5137/1019-5149.JTN.27986-19.3	11.1 %	10.0 (3.1)	NA
Microthrombi Correlates With Infarction and Delayed Neurological Deficits After Subarachnoid Hemorrhage in Mice	Dienel (2020)	10.1161/STROKEAHA.120.029753	13.3 %	NA	90.4 % (9.2 %)
Minimal Long-Term Neurobehavioral Impairments after Endovascular Perforation Subarachnoid Hemorrhage in Mice	Fanizzi (2017)	10.1038/s41598-017-07701-y	28.0 %	NA	NA
Morphological Characteristics of Neuronal Death After Experimental Subarachnoid Hemorrhage in Mice Using Double Immunoenzymatic Technique	Nakano (2019)	10.1369/0022155419878181	63.4 %	6.8 (2)	NA
MRI-based in vivo assessment of early cerebral infarction in a mouse filament perforation model of subarachnoid hemorrhage	Sasaki (2017)	10.1016/j.neuret.2017.05.047	4.0 %	NA	NA

Title	Author and year	DOI	Mortality	SAH grade	Vasospasm
New grading system based on magnetic resonance imaging in a mouse model of subarachnoid hemorrhage	Egashira (2015)	10.1161/STROKEAHA.114.007834	19.0 %	NA	NA
Oxidative stress after subarachnoid hemorrhage in gp91phox knockout mice	Liu (2007)	10.1017/s0317167100000682x	10.9 %	NA	NA
Role of the endothelium NO-Synthase in early brain injury after experimental subarachnoid hemorrhage	Lenz (2017)	10.1177/0271678X17695982	0.0 %	NA	NA
Single clip: An improvement of the filament-perforation mouse subarachnoid haemorrhage model	Peng (2019)	10.1080/02699052.2018.1531310	25.0 %	10.2 (5.5)	NA
Standardized induction of subarachnoid hemorrhage in mice by intracranial pressure monitoring	Feiler (2010)	10.1016/j.jneumeth.2010.05.005	30.0 %	NA	NA
Stimulator of IFN genes mediates neuroinflammatory injury by suppressing AMPK signal in experimental subarachnoid hemorrhage	Peng (2020)	10.1186/s12974-020-01830-4	17.9 %	11.0 (1.5)	NA
Subarachnoid hemorrhage in C57BL/6J mice increases motor stereotypies and compulsive-like behaviors	Nanegrungsunk (2021)	10.1080/01616412.2020.1841481	33.3 %	NA	NA
TSPO ligand Ro5-4864 modulates microglia/macrophages polarization after subarachnoid hemorrhage in mice	Zhou (2020)	10.1016/j.neulet.2020.134977	14.6 %	10.4 (1.5)	NA
Ultra-Early Cerebral Thrombosis Formation After Experimental Subarachnoid Hemorrhage Detected on T2* Magnetic Resonance Imaging	Wang (2021)	10.1161/STROKEAHA.120.032397	0.0 %	NA	NA
Value of Three-Dimensional Maximum Intensity Projection Display to Assist in Magnetic Resonance Imaging (MRI)-Based Grading in a Mouse Model of Subarachnoid Hemorrhage	Mutoh (2016)	10.12659/msm.896499	18.9 %	NA	NA
White Matter Injury After Subarachnoid Hemorrhage: Role of Blood-Brain Barrier Disruption and Matrix Metalloproteinase-9	Egashira (2015)	10.1161/STROKEAHA.115.010351	22.2 %	9.0 (3.3)	NA
White matter T2 hyperintensities and blood-	Toyota (2019)	10.1111/cns.13221	0.0 %	NA	NA

Title	Author and year	DOI	Mortality	SAH grade	Vasospasm
brain barrier disruption in the hyperacute stage of subarachnoid hemorrhage in male mice: The role of lipocalin-2					
Sevoflurane and Desflurane Exposures Following Aneurysmal Subarachnoid Hemorrhage Confer Multifaceted Protection against Delayed Cerebral Ischemia	Jayaraman (2021)	10.3390/biomedicines9070820	NA	NA	74.2 % (19.5 %)

Table of included articles in meta-analysis. Articles had to report at least one of mortality, SAH grade and vasospasm. Mortality is presented as the share of animals that died after SAH perforation induction. We list SAH grades exclusively that were in accordance with the scoring system proposed by Sugawara et al. (167). Vasospasm is presented as the relative MCA, ACA, ICA and basilar artery diameter after SAH perforation induction in comparison to corresponding sham-operated mice artery diameter. Values in brackets represent standard deviations. NA = not applicable (means that the respective outcome was not reported).

7.3 Questionnaire for missing experimental parameter information in publication 1

Title of the questionnaire: Systematic Review and Meta-Analysis of reporting on quality control standards and phenotypes of the basic cell model in brain cancer research: how reproducible is the field?

Instruction for the authors:

Please answer the following questions as accurately as possible. All questions refer to the experiments investigating the effect of Temozolomide monotherapy on U87-MG cell viability.

Description (not visible for the authors):

The questionnaire was sent to all authors for whom at least one of the following experimental parameters was not or not clearly reported in the respective original publication included in the systematic review. The questionnaire was first sent to the correspondence email address indicated in the publication, which was repeated once if no response was received after two weeks. Also, if there was still no response or no correspondence email address was indicated, research was carried out to find a different email address for the corresponding and first author who were also contacted. Only those questions for which the requested parameters were not reported were asked, so that an individual questionnaire was created for each publication. In cases where multiple included publications were assigned to a corresponding author, the author was sent a separate questionnaire for each publication. All responses were used anonymously in the recorded statistics.

Questionnaire:

Question 1 (Q1): How is the glucose-level of the U87-MG cell culture medium DMEM?

- high-glucose
- low-glucose
- no glucose
- other:

Q2: Which type of untreated control did you use for the comparison of the effect of Temozolomide on the viability/proliferation of U87-MG cells?

- medium only
- medium and Phosphate-buffered saline (PBS)
- medium with drug vehicle (Dimethyl sulfoxide (DMSO))
- other:

Q3: Was a genomic authentification of the U87-MG cell line performed?

- Yes
- No
- other:

Q4: Did you make any restrictions for using U87-MG cells in your experiments in terms of the age of the cells (e.g. based on time or a maximum cell passage)? If applicable, please describe the limitation. If you did not make such limitations, please also note it.
-free text field-

Q5: At what concentration of Temozolomide were the U87-MG cells incubated?
-free text field-

Q6.1: Was the medium changed during the incubation of U87-MG cells with Temozolomide?

- yes, the medium was changed during the time of incubation
- no, the medium was not changed during the time of incubation
- other:

Q6.2: If the medium was changed during the time of incubation of U87-MG cells with Temozolomide, in which way did you change it?
-free text field-

Q7: How long were the U87-MG cells incubated with Temozolomide?
-free text field-

Q8.1: Was the cell proliferation/viability measured directly after the incubation of U87-MG cells with Temozolomide or was there a break in time between the Temozolomide incubation and the measurement?

- The U87-MG cell viability/proliferation was measured directly after the incubation with Temozolomide.
- There was a break in time between the Temozolomide incubation of U87-MG cells and the cell viability/proliferation measurement.
- Other:

Q8.2: If there was a break in time between the Temozolomide incubation and the U87-MG cell viability/proliferation measurement, how long was the break?
-free text field-

Q9.1: What is the number of independent experiments that you performed to obtain the presented cell proliferation data of U87-MG cells exposed to Temozolomide?
-free text field-

Q9.2: What is the number of replications per independent experiment that you performed in the cell proliferation experiments with U87-MG cells exposed to Temozolomide?
-free text field-

Q10: What is the errortype of the presented proliferation inhibition results of U87-MG cells exposed to Temozolomide monotherapy?

- Standard Deviation (SD)
- Standard Error of the Mean (SEM)
- Other:

Q11: How did you calculate your presented cell proliferation inhibition results of the experiments with U87-MG cells exposed to Temozolomide?

- average and error of every single cell viability value measured from every independent experiment and each replication
- average and error of one independent experiment are shown as representative data
- calculation of the average and error of the calculated averages and errors of each independent experiment to get a summarised data point for all independent experiments
- other:

Acknowledgements

I am very grateful for the supervision of my research by Doctor Sajjad Muhammad, who provided advice and guidance at every stage of my work. I would also like to thank Professor Malcolm Macleod who, despite the geographical distance, was always able to give me prompt feedback on the project. Moreover, I am thankful to Professor Ulf Kahlert for his valuable advice on the realisation of the project. Furthermore, I appreciate the constructive comments of the entire research team led by Doctor Sajjad Muhammad, with special thanks to Doctor Serdar Alpdogan for his support. I also acknowledge the CAMARADES (Collaborative Approach to Meta Analysis and Review of Animal Experimental Studies) research group in Edinburgh for their valuable advice on the conduct of the meta-reviews, with special thanks to Emma Wilson for her prompt correspondence. I would also like to take this opportunity to thank the experienced statisticians Doctor Igor Fischer, Professor Holger Schwender and Professor Shinichi Nakagawa for critically reviewing my analytical methods. Last but not least, I express my deep gratitude to my family, my friends, and my partner for their constant support in the course of this research project.