








Meta-analysis on reporting practices as a source of heterogeneity in in vitro cancer research

Timo Sander ,¹ Joly Ghanawi ,² Emma Wilson ,³ Sajjad Muhammad ,¹ Malcolm Macleod ,³ Ulf Dietrich Kahlert ⁴

This article has received OSF badges for Open data, Open materials and Pre-registration.

To cite: Sander T, Ghanawi J, Wilson E, *et al*. Meta-analysis on reporting practices as a source of heterogeneity in in vitro cancer research. *BMJ Open Science* 2022;**6**:e100272. doi:10.1136/bmjos-2021-100272

► Prepublication history and additional supplemental material for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjos-2021-100272>)

Received 14 December 2021
Accepted 09 May 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

¹Department of Neurosurgery, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany

²Independent researcher, Edinburgh, UK

³Centre for Clinical Brain Sciences, The University of Edinburgh Medical School, Edinburgh, UK

⁴Department of Molecular and Experimental Surgery, Clinic for General, Visceral, Vascular and Transplant Surgery, Otto von Guericke Universität Magdeburg, Magdeburg, Germany

Correspondence to

Professor Ulf Dietrich Kahlert; ulf.kahlert@med.ovgu.de

ABSTRACT

Objectives Heterogeneity of results of exact same research experiments oppose a significant socioeconomic burden. Insufficient methodological reporting is likely to be one of the contributors to results heterogeneity; however, little knowledge on reporting habits of in vitro cancer research and their effects on results reproducibility is available. Exemplified by a commonly performed in vitro assay, we aim to fill this knowledge gap and to derive recommendations necessary for reproducible, robust and translational preclinical science.

Methods Here, we use systematic review to describe reporting practices in in vitro glioblastoma research using the Uppsala-87 Malignant Glioma (U-87 MG) cell line and perform multilevel random-effects meta-analysis followed by meta-regression to explore sources of heterogeneity within that literature, and any associations between reporting characteristics and reported findings. Literature that includes experiments measuring the effect of temozolomide on the viability of U-87 MG cells is searched on three databases (Embase, PubMed and Web of Science).

Results In 137 identified articles, the methodological reporting is incomplete, for example, medium glucose level and cell density are reported in only 21.2% and 16.8% of the articles. After adjustments for different drug concentrations and treatment durations, the results heterogeneity across the studies ($I^2=68.5\%$) is concerning large. Differences in culture medium glucose level are a driver of this heterogeneity. However, infrequent reporting of most experimental parameters limits the analysis of reproducibility moderating parameters.

Conclusions Our results further support the ongoing efforts of establishing consensus reporting practices to elevate durability of results. By doing so, this work can raise awareness of how stricter reporting may help to improve the frequency of successful translation of preclinical results into human application. The authors received no specific funding for this work. A preregistered protocol is available at the Open Science Framework (<https://osf.io/9k3dq>).

INTRODUCTION

Progress in scientific research is a dynamic process which thrives in the interaction of diverse research groups addressing shared problems. The scientific model has new

STRENGTHS AND LIMITATIONS OF THE STUDY

- ⇒ This is a novel systematic review and meta-analysis focusing on the reproducibility of a commonly performed in vitro drug test assay in cancer research, including a large sample size of evidence with 137 articles.
- ⇒ We used a three-level meta-analytical random-effects model accounting for dependence of data from the same article.
- ⇒ Full availability of the statistical code used and the underlying data is offered for transparent methods.
- ⇒ The statistical power in meta-regressions investigating heterogeneity moderators is limited due to infrequent reporting of experimental parameters.
- ⇒ We experienced a low rate of response (8.0%) when asking the authors of the included articles for additional information regarding unreported parameters.

findings either confirmed or refuted by other scientists, so that science becomes self-correcting.¹ However, one key foundation for self-correction is that key experimental methods needed for interpretation and repetition of published research are described in sufficient detail. Recent efforts to replicate key findings in cancer biology and further research areas have raised scientific, ethical and economic concerns.²⁻⁹

Glioblastoma is a malignant brain tumour with a median time of survival of around 15 months.¹⁰ First-line treatment consists of a multimodal approach of surgical resection followed by radiation therapy and chemotherapy with the alkylating agent temozolomide (TMZ).¹¹ A previous systematic review of the in vivo literature describing the efficacy of TMZ showed limited reporting of key study design features and low prevalence of reporting of measures to reduce risks of bias.¹² In vitro glioblastoma research commonly uses the commercially available cell line Uppsala-87 Malignant Glioma (U-87 MG),^{13 14} originally derived in 1966 from a female patient at Uppsala University.¹⁵ However, the currently



available U-87 MG line distributed by the American Type Culture Collection (ATCC, Manassas, Virginia, USA)¹⁶ has been found to be different from the original version.¹⁷ It is unclear to what extent these U-87 MG cells are truly representative of the original tumour tissue and whether they allow for reproducible experiments when serving as glioblastoma models.

The reproducibility of in vitro glioma research is likely to depend on the completeness of reporting of key study design features, including reporting of risks of bias. Here, we use a systematic review to portray the in vitro literature describing the effectiveness of TMZ in reducing the growth of U-87 MG cells, with a focus on the use of clinically relevant drug concentrations and treatment durations, on methodological reporting and how this might influence the reproducibility of results.

METHODS

A preregistered study protocol is available at the Open Science Framework¹⁸ and was uploaded before full text based screening and data extraction began. Deviations from the protocol are described at the appropriate sections.

SYSTEMATIC REVIEW

Systematic literature search and screening

The systematic search was conducted on the databases PubMed, Embase and Web of Science on 26 August 2020 using the search strategy described in the online supplemental additional data A1.¹⁹ Two reviewers independently screened article titles and abstracts for potential inclusion, with discrepancies resolved by a third reviewer. This was followed by full text screening. We included studies describing controlled in vitro cell culture experiments that compared the effect of a single TMZ treatment on the viability of U-87 MG cells with that in untreated controls. We also required that cell viability was measured by colorimetric assay or by cell counting, and that the authors used Dulbecco's Modified Eagle Medium (DMEM) as the ATCC recommends using a modified Eagle medium for U-87 MG cell cultures¹⁶ and as a recent review identified DMEM as the most frequently used medium for this type of experiment.¹⁴ This culture medium restriction was introduced to guarantee sufficient comparability of experiments across studies and to prevent interference of different culture media in the analysis of the impact of culture medium ingredients (eg, glucose concentration and antibiotics) on the overall effect and its reproducibility. We only included original peer-reviewed research articles in the English language, with no restrictions made on the publication year. Our protocol had included consideration of cell growth rates in xenotransplantation models, but we later decided to focus exclusively on in vitro research. Detailed inclusion and exclusion criteria are given in online supplemental additional data A2.¹⁹

Data extraction

We recorded effect sizes for change in cell viability in response to TMZ compared with untreated control, TMZ concentration and duration of exposure. As shown in online supplemental additional data A3 and A4,¹⁹ we recorded 18 experimental parameters, 8 risks of bias items and the journal impact factor (JIF) for that journal in the year of publication.²⁰ Consideration of JIF had not been included in our study protocol and should be considered an exploratory analysis.

Two reviewers independently recorded study design features, risk of bias items and effect sizes, with reconciliation of discrepancies by a third reviewer where necessary; data for effect sizes were reconciled if they differed by more than 15% of the largest effect size; otherwise, the mean of the extracted values from the two reviewers was used. Where information was missing, we contacted authors for clarification; to reduce the burden on them to reply, we limited this request to a maximum of 11 items as described in the online supplemental additional data A5.¹⁹

Reporting quality of experimental parameters and articles

Where an article reported multiple relevant experiments, a parameter was considered as reported for an article if it was reported in every belonging experiment. When calculating the number of items reported for each article, we did not consider the volume of added TMZ and of control fluid as we considered this was included if the drug and control concentration was given. We analysed change in reporting quality over time, and any relationship between reporting quality and JIF, using linear regression.

META-ANALYSIS

Exclusions from the meta-analysis

We excluded experiments that did not report data essential for our analysis such as cell viability data for the untreated control group, TMZ concentration, duration of treatment or the number of experimental units; and we excluded baseline data (where treatment duration=0).

Effect size

We calculated a cell viability reduction caused by TMZ compared with the corresponding untreated control as raw mean difference with all data given in relation to the control viability as:

$$\text{Cell_viability_reduction} = D = \frac{\text{Viability}_{\text{Con}} - \text{Viability}_{\text{TMZ}}}{\text{Viability}_{\text{Con}}}$$

with its variance:

$$V_D = \frac{n_{\text{Con}} + n_{\text{TMZ}}}{n_{\text{Con}} \times n_{\text{TMZ}}} \times \frac{(n_{\text{Con}} - 1) \left(\frac{SD_{\text{Con}}}{\text{Viability}_{\text{Con}}} \right)^2 + (n_{\text{TMZ}} - 1) \left(\frac{SD_{\text{TMZ}}}{\text{Viability}_{\text{Con}}} \right)^2}{n_{\text{Con}} + n_{\text{TMZ}} - 2}$$

where n_{Con} and n_{TMZ} represented the number of experiments in the control and TMZ group, respectively, and SD_{Con} and SD_{TMZ} represented the SD of cell viabilities in the control and TMZ group. If the variance in the control group was not reported, we assumed this to be equivalent to variance in the corresponding TMZ group. If it was not

clear whether variance was reported as SD or SEM, we assumed they were SEM, as a more conservative approach.

Multilevel random-effects meta-analysis

We used random-effects meta-analysis.²¹ Because a single article might contribute several effects sizes, we used a three-level model, where the first level represented the raw cell viability data, the second level represented all the effects from a given article and the third level represented the article itself. This accounts for the relative non-independence of effects reported in the same article.²² Moreover, as the exact correlations of the dependent effects within an article were unknown, we used robust variance estimation.²³

To estimate τ^2 , we used the restricted-maximum-likelihood method.²⁴ This method has recently been shown to be robust for non-normally distributed effect sizes²⁵ and is recommended for the estimation of τ^2 .^{25, 26} We used the t-distribution for the calculation of the weighted mean effect as this accounts for uncertainty in the estimation of τ^2 .²⁷ We took the within-level-three estimate of τ^2 as a measure of reproducibility of findings between studies and subsequently as an indicator of irreproducibility of results across articles.

Meta-regression

We tested 10 a priori defined parameters of which we expected reproducibility moderation (plus article reporting, TMZ concentration and treatment duration) in univariable meta-regressions. For these, a reduction of within-level-three τ^2 would indicate that the parameter moderated reproducibility, with lower within-level-three τ^2 indicating that findings would be more likely to reproduce if that parameter was controlled between the original and replicating experiments. We took the same approach to establish any effect of articles overall reporting quality or JIF. We also used univariable meta-regression to analyse the effects of TMZ dose and treatment duration. We transformed TMZ concentrations into a four-parameter log-logistic dose-response model.²⁸

A similar four-parameter log-logistic time-response model was used for the treatment durations. Finally, we conducted multivariable meta-regression of the effect of dose, duration and moderators proven significant in univariable meta-regression. In all analysis, we set a significance level of 0.05.

Analysis of reporting of published work since the time point of systematic search

Due to the relatively long interval between systematic literature search and publication of this review, we conducted an updated literature search on 10 March 2022 to investigate whether reporting quality may have changed during this period. Thereby, we assessed the study parameter reporting and the prevalence of potential risks of bias in the identified additional articles. The same rules for literature screening and data collection were applied as described above. Changes in overall reporting quality

and risks of bias prevalence were tested with the Mann-Whitney-Wilcoxon test.

Software

To remove duplicate articles from the systematic search results, we used two approaches, the deduplication function integrated in Zotero²⁹ and one developed by the CAMARADES group.³⁰ Articles were removed if they were detected as a duplicate by both functions. Afterwards, articles were manually screened for remaining duplicates. Screening and data extraction used the Systematic Review Facility for preclinical systematic reviews.³¹ Graphically presented data were extracted with the WebPlotDigitizer.³² Meta-analysis was performed within the RStudio environment³³ using the programming language R.³⁴ We used the `rma.mv` function of the R package `Metafor`³⁵ for multi-level meta-regressions, the R package `clubSandwich`³⁶ for robust variance estimation, the R package `orchard`³⁷ for the calculation of marginal R^2 and I^2 , the `drm` function of the R package `drc`³⁸ for the dose-response and time-response models and the `lm` function of the integrated R package `stats`³⁴ for linear regressions. The full statistical code and its datasets are publicly available on the GitHub data repository.³⁹

RESULTS

Systematic search

We identified 1158 publications of which 137 articles met our inclusion criteria and were included in the systematic review; 101 provided sufficient data to be included in the meta-analysis (figure 1). Online supplemental additional data A6¹⁹ contains a list of all included articles. These 137 articles described 828 experiments, where every different combination of drug concentration and treatment duration used was considered as an individual experiment. The main reason for exclusion from meta-analysis was an unreported number of contributing experimental units (n=24 articles).

Experimental parameters distribution

Across 137 publications, a broad range of experimental characteristics was observed. The most common source of U-87 MG cells was the ATCC (66 articles, 48.2% of all included articles; table 1). A cell line authentication report was available in 16 articles (11.7%), and 8 articles (5.8%) described testing for mycoplasma contamination. The reported cell passage number ranged from 3 to 100 (median of 15), but 123 publications (89.8%) did not report it. Only 29 of 137 articles (21.2%) reported the level of glucose in culture medium, and in these high glucose supplementation (4500 mg/dL) was most prevalent (in 24 of 29). Control treatment was dimethyl sulfoxide (DMSO) in 37 articles and culture medium alone in 13 articles; in 87 articles, it was not reported or only labelled as 'untreated control' without further specification. The most common cell viability assessment method was the 3-(4,5-dimethylthiazol-2-yl)-2,5-dipheny

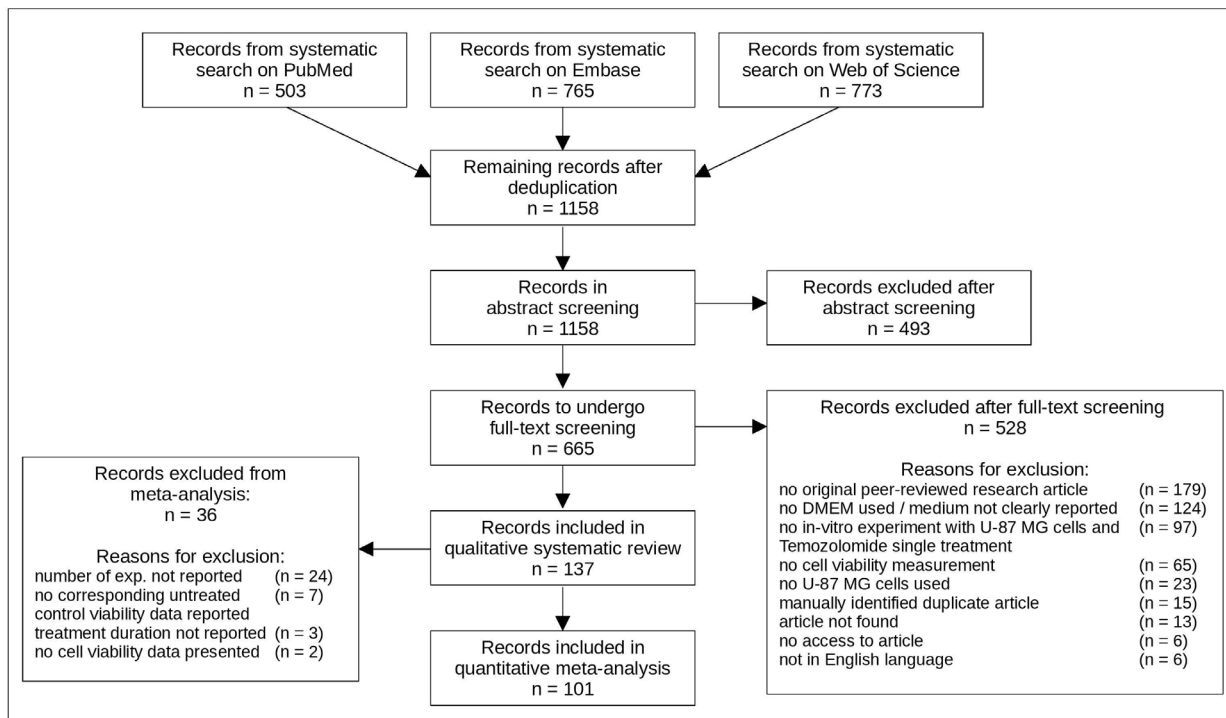


Figure 1 Systematic search and screening results. Presentation based on the PRISMA statement.⁶⁸ Systematic search was conducted in August 2020. Qualitative analysis included all calculations in this review except meta-analysis and meta-regressions. One reason for exclusion per excluded article. DMEM, Dulbecco's Modified Eagle's Medium; Exp, experiments; U-87 MG, Uppsala-87 Malignant Glioma.

Itetrazolium bromide (MTT) assay (67 articles, 48.9%). The concentration of U-87 MG cells used ranged from 5 cells/ μL to 500 cells/ μL (table 2), with a median of 30 cells/ μL . Ninety-three articles included information for the number of cells per well but not the volume in which these cells were plated. Cells were passaged based on confluence (13 articles, ranging from 50% to 90% confluence) or on time (4 articles, ranging from 2 days to 7 days), but criteria for cell passaging were not stated in 120 studies (87.6%; table 3).

Overall, 98 different TMZ concentrations (10nM–16.0mM; median: 100 μM) and 20 different treatment durations (4 hours–12 days; median: 3 days) were reported, as illustrated in online supplemental additional data A7 and A8, respectively. In several articles, it was not clear whether cell viability was measured directly after TMZ exposure, or whether there was a 'wash out' or recovery period. In others, it was unclear whether TMZ was added to the cells once and remained in suspension or whether TMZ was added repeatedly at different times. For the purposes of the meta-analysis, we assumed a single TMZ addition with continuous incubation for the reported time followed directly by the assessment of cell viability.

Completeness of reporting

Several key experimental parameters were reported in fewer than half of the articles: The type of untreated control was reported in 36.5%, the glucose level of culture mediums in 21.2% and U-87 MG cell age in 7.3% of all

137 articles (figure 2A). The median number of quality items reported was 8.4 of 16 (ranging from 3 to 13). Analysis of change over time suggested slight reporting improvement (+0.635% parameter reporting ratio per article per year, $p=0.011$; figure 2B); and reporting quality seemed to be higher for articles published in journals with higher impact factors (+1.74% per unit increase in JIF unit, $p<0.001$; figure 2C).

Reporting of measures to reduce risks of bias

Not one of 137 articles described a sample size calculation, random allocation to experimental group, blinded outcome assessment or the use of a preregistered protocol specifying the hypotheses and outcomes as shown in the online supplemental additional data A9.¹⁹ The methods used to calculate cell viability average and error values were unclear in 92 articles, and the number of independent experiments and technical replicates per experiment conducted were unclear in 47 articles. The mean number of measures to reduce risks of bias reported was 2.9.

Reporting and risks of bias in publications published after systematic search conduction

In the interval between the initial systematic literature search (August 2020) and March 2022, the literature flow chart shown in online supplemental additional data A10¹⁹ presents 41 additionally published articles. As shown in online supplemental additional data A11,¹⁹ most experimental parameters offered negligible variations between

Table 1 Extracted parameters

Parameter	Phenotype	Articles	
General article information			
Conflict of interests statement	Declaration of no conflict of interests	86	62.8%
	Declaration of existing conflict of interests	5	3.6%
	No statement about conflict of interests	46	33.6%
U-87 MG in vitro model			
Source of U-87 MG cells	ATCC, Manassas, Virginia, USA	66	48.2%
	Chinese Academy of Sciences, Beijing, China	27	19.7%
	Other commercial/institutional sources	24	17.5%
	Colleagues	11	8.0%
	Not reported	9	6.6%
U-87 MG cell line authentication conducted	Yes	16	11.7%
	No/not reported	121	88.3%
U-87 MG age (maximum number of cell passage)	3	1	0.7%
	7	1	0.7%
	8	1	0.7%
	10	3	2.2%
	15	4	2.9%
	20	2	1.5%
	35	1	0.7%
	100	1	0.7%
	Not reported	123	89.8%
U-87 MG culture conditions			
Glucose level of cell culture medium	Low glucose (1000 mg/dL)	3	2.2%
	High glucose (4500 mg/dL)	24	16.8%
	Low and high glucose (in different experiments)	1	0.7%
	Without glucose	1	0.7%
	Not reported	108	78.8%
Mycoplasma contamination checked	Yes	8	5.8%
	Not reported	129	94.2%
Supplemented antibiotics	Penicillin and streptomycin	92	67.2%
	Other antibiotics	5	3.6%
	No antibiotics supplemented	3	2.2%
	Not reported	37	27.0%
Source of FBS	Thermo Fisher Scientific, Waltham, Massachusetts, USA (including Gibco, Invitrogen and Life Technologies)	51	37.2%
	Hyclone Laboratories, Logan, Utah, USA	13	9.5%
	Sigma-Aldrich, St Louis, Missouri, USA	8	5.8%
	Other sources	22	16.1%
	FBS was not used	1	0.7%
	Not reported	42	30.7%
Control group and outcome measurement			
Type of untreated control	Drug vehicle (DMSO)	37	27.0%
	Cell culture medium only	13	9.5%
	Not reported	87	63.5%

Continued

Table 1 Continued

Parameter	Phenotype	Articles	
Cell viability assessment method	MTT assay, colorimetric	67	48.9%
	CCK8, colorimetric	20	14.6%
	SRB assay, colorimetric	9	6.6%
	Alamar Blue assay, colorimetric	7	5.1%
	Trypan Blue Exclusion test, cell counting	6	4.4%
	WST-1 assay, colorimetric	6	4.4%
	MTS assay, colorimetric	3	2.2%
	Other assessment methods	11	8.0%
	More than one assay used	8	5.8%

Extracted parameter phenotypes (including additional information obtained through contacting the authors). One phenotype per article. The column 'Articles' shows the absolute and relative frequencies of articles with the parameter phenotype in relation to all 137 included articles.

ATCC, American Type Culture Collection; CCK8, Cell Counting Kit-8; DMEM, Dulbecco's Modified Eagle's Medium; DMSO, dimethyl sulfoxide; FBS, fetal bovine serum; U-87 MG, Uppsala-87 Malignant Glioma; MTS, 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; SRB, Sulforhodamine B; WST-1, water soluble tetrazolium 1.

the time periods, including the glucose level of cell culture medium (−1.7% less reported compared with initial search). The largest improvements comprised conflicts of interests (+26.3 %) and cell concentration reporting (+17.3 %). One notable deterioration was less frequent explicit reporting of independent experiments and their technical replicates (−21.8 %), resulting in a lower rate of unambiguous computational pathways for cell viability mean and error values (−13.3 %). However, the only significant change was observed in an improved conflicts of interests reporting. A detailed comparison of reporting and risks of bias is given in the online supplemental additional data A12.¹⁹ Meanwhile, the overall

articles reporting quality and prevalence of risks of bias has not changed significantly as indicated in online supplemental additional data A13 and A14¹⁹ ($p=0.390$ and $p=0.751$, respectively (Mann-Whitney-Wilcoxon test)).

Meta-analysis of the effect of TMZ

The observed effect of TMZ is highly heterogeneous; variation within different experiments in the same article (represented by I^2 of level-two variance) accounts for 56.6% of observed variance, variation of the effect across different articles (represented by I^2 of level-three variance) accounts for 42.9% of observed variance, and the variance due to random chance expected if all experiments of all articles were held under identical conditions made up only 0.5% of the total variance (table 4). The

Table 2 Extracted cell concentrations

Cell concentration (cells/ μ L)	Articles	
5	1	0.7%
12.5–62.5	1	0.7%
15	1	0.7%
20	3	2.2%
25	3	2.2%
30	4	2.9%
40	2	1.5%
50	5	3.6%
100	1	0.7%
166.7	1	0.7%
200	1	0.7%
500	1	0.7%
Reporting of only the number of cells per well without the associated volume per well	93	67.9%
No information regarding the cell number, the volume they are plated in or the cell concentration was given	20	14.6%

Extracted cell concentrations (including additional information obtained through contacting the authors). One concentration per article. The column 'Articles' shows the absolute and relative frequencies of articles with the parameter phenotype in relation to all 137 included articles.

Table 3 Extracted cell passaging criteria

Criterion	Articles	
Based on cell culture confluence		
50%–70%	1	0.7%
60%–80%	1	0.7%
70%	2	1.5%
70%–90%	3	2.2%
80%	6	4.4%
Based on time intervals		
2 days	1	0.7%
2–3 days	1	0.7%
3–4 days	1	0.7%
7 days	1	0.7%
No cell passaging criteria were reported	120	87.6%

Extracted cell passaging criteria (including additional information obtained through contacting the authors). One criterion per article. The column 'Articles' shows the absolute and relative frequencies of articles with the parameter phenotype in relation to all 137 included articles.

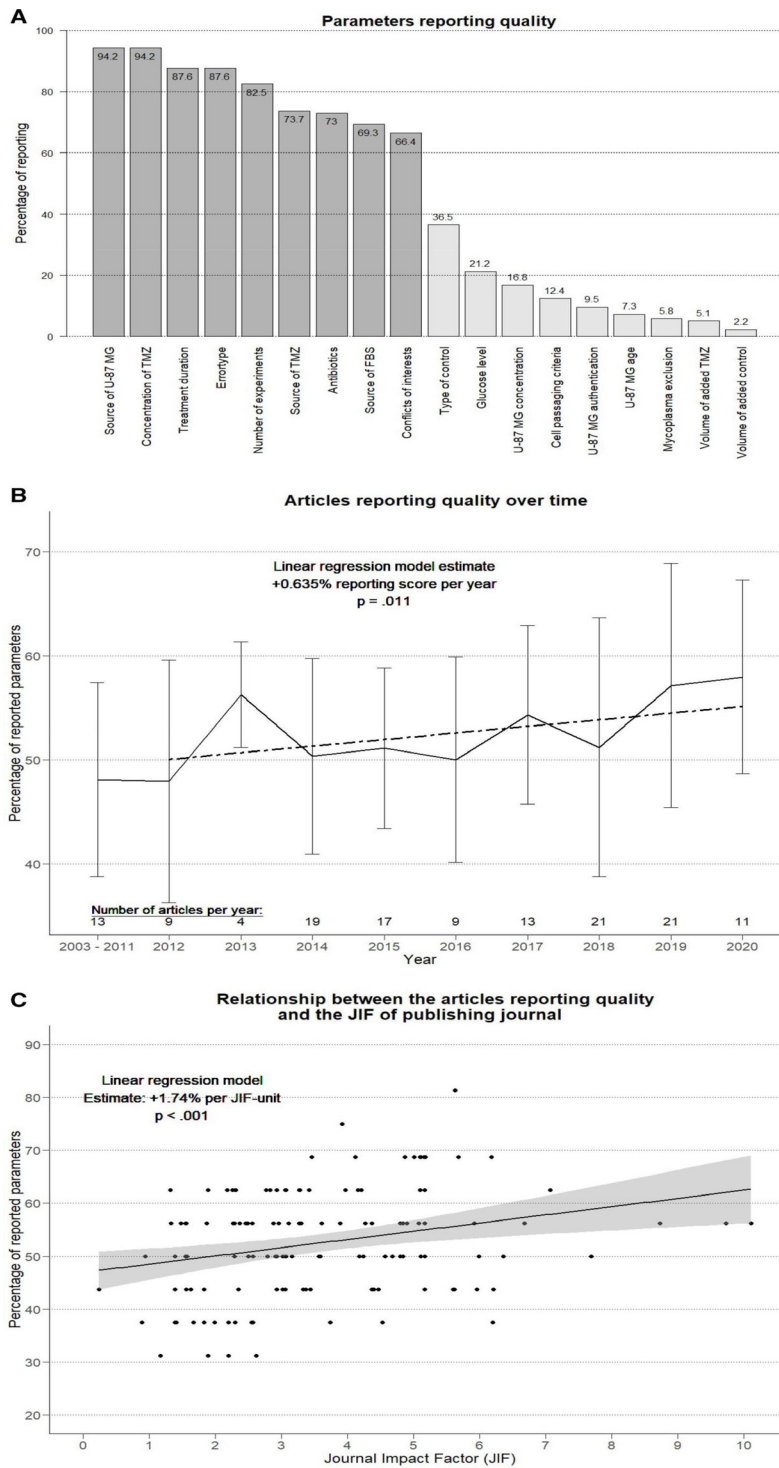


Figure 2 Reporting quality (of parameters, over time and depending on the JIF). (A) The reporting quality of a parameter was defined as the share of articles that reported this parameter in comparison to all 137 included articles. This share is shown on the top of each bar for each parameter. A parameter was considered to be reported if its phenotype was clear based on the information provided in the original full-text research article. (B) Linear regression model of articles reporting quality (proportion of reported parameters to 16 experimental parameters) and year of publication. The articles published before 2012 were graphically summarised because of low numbers of articles published in these years (but the exact years were used for the regression). Only articles published until the time of systematic search in August 2020 were considered. The dotted line represents the linear regression prediction; error bars indicate the SD around the mean reporting score per year represented by the continuous line. (C) Linear regression model of articles reporting quality and the JIF of the articles publishing journal in the year of publication. The grey area marks the 95% CI of the regression model prediction. JIFs were obtained from the Clarivates Incites Journal Citation Reports (Web of Science group, 2020). For the articles published in 2020, the JIF of 2019 replaced the JIF of 2020 as the more recent was not available at the time of analysis. One article was omitted as no JIF could be obtained. FBS, fetal bovine serum; JIF, journal impact factor; TMZ, temozolomide; U-87 MG, Uppsala 87 Malignant Glioma.

Table 4 Random-effects three-level meta-analysis suggests significant irreproducibility

Meta-analysis data	
Number of included effects of TMZ on U-87 MG viability	644
Number of included articles the effects were reported in	101
Effect of TMZ	
Overall weighted mean effect of TMZ (U-87 MG viability reduction compared to untreated control)	33.8% (30.0% to 37.7%)
Investigation of heterogeneity	
Test whether heterogeneity is present	
Q (df=643)	134 066.1
P value	<0.001
Total I ²	99.5%
Within-articles variance of the true effect	
τ^2	3.6% (3.2% to 4.1%)
τ	19.0% (17.9% to 20.3%)
I ² (=proportion of total variance)	56.6%
Between-articles variance of the true effect (~irreproducibility)	
τ^2	2.8% (1.9% to 3.9%)
τ	16.6% (13.9% to 19.8%)
I ² (=proportion of total variance)	42.9%

Random-effects three-level meta-analysis using the raw data the effects were calculated with as first level, the effect sizes within each article as second level and the articles the effects were reported in as third level. τ^2 : estimator of the variance of true effects (level-two variance: within-articles variance; level-three variance: between-articles variance: representant of irreproducibility); τ : square root of τ^2 ; I²: proportion of within-articles and between-articles variance, respectively, of the total observed variance, including sampling error. τ^2 estimator: restricted-maximum likelihood. Cochran's Q was used as the test for heterogeneity using a χ^2 distribution; values in parenthesis show CIs with significance level set at 0.05. df, degree of freedom; U-87 MG, Uppsala-87 Malignant Glioma; TMZ, temozolomide.

heterogeneity of results across the articles is reflected in an SD (τ) of $\pm 16.6\%$ (for this SD estimate, 95% CI 13.9% to 19.8%) around a global estimate of a reduction in cell viability following TMZ treatment of 33.8% (95% CI 30.0% to 37.7%) compared to the untreated control.

Drivers of heterogeneity

The within-articles variance of effects reported in the same article was, as expected, partly explained by differences in TMZ concentrations (48.3%) and treatment durations (4.8%) and is presented in the online supplemental additional data A15.¹⁹ However, both features did not explain parts of between-articles variance of effects (table 5). Combining TMZ concentration and treatment duration in a multivariable meta-regression reduced within-articles variance τ^2 from 3.6% to 1.7%, while the

estimated between-articles variance τ^2 increased from 2.8% to 3.7% (table 6).

The glucose level in the cell culture medium was the only experimental parameter significantly associated with heterogeneity of results across the articles ($p=0.016$, table 5). The moderator fit indicated by marginal R² was 7.0%, and 10.9% of the between-articles variance τ^2 could be explained because of differences in the glucose level. In other words, roughly 11% of results variation was attributed to different glucose levels used in the culture medium. The different glucose levels and effects are shown in online supplemental additional data A16,¹⁹ indicating a smaller effect of TMZ in the articles using high glucose supplementation (cell viability reduction of 23.1% compared to the untreated control, 95% CI 15.8% to 30.5%) than in the articles using an unreported glucose level (37.1%, 95% CI 28.6% to 45.6%, $p=0.002$).

Articles reporting quality showed a significant correlation with the reported effect of TMZ (marginal R²=3.3%). Reported effect on cell viability fell by 3.0% for each unit increase in the number of reported parameters ($p=0.026$). Adding these two features (glucose level and articles reporting quality) to the multivariable between-articles variance meta-regression, including drug concentration and treatment duration, resulted in a slightly improved model, reducing τ^2 from 3.7% to 3.5%.

DISCUSSION

Reporting and experimental parameters

We found a highly significant relationship between the concentration of TMZ used, the duration of treatment and the measured effect of TMZ on the viability of U-87 MG cells ($p<0.001$). However, the reporting of experimental parameters in this literature—including such fundamental issues as the concentration of the drug and the treatment duration—is limited. A recent study has also identified suboptimal reporting of basic experimental parameters and varying cell viability reducing effects of TMZ in in vitro glioma cell line experiments with TMZ single treatment.¹⁴

Based on TMZ concentrations found in peritumoral tissue^{40 41} and cerebrospinal fluid,⁴² Stepanenko and Checkhonin recommended in vitro studies should use TMZ concentrations of 1–10 μM .⁴³ Although the effect of higher concentrations could be examined, it seems reasonable to expect that publications use at least one clinically relevant drug concentration. More than two-thirds of articles (70 of 101 articles included in meta-analysis) did not use clinically relevant TMZ concentrations in at least one of their experiments. The effects of TMZ are due to DNA alkylation and methylation, and this requires TMZ internalisation, which usually occurs during cell division. An effect of TMZ can, therefore, be expected at the earliest after about 1.5 cell doubling times, and the cell doubling time of U-87 MG is around 34 hours.^{44 45} Applying an early credible limit for efficacy of 51 hours, 31.7% of the articles included in the meta-analysis only measured

Table 5 Moderators of between-articles variance of true effects

Moderator	Type	Number of effects	Number of articles	P value	Marginal R ²	Between-articles variance		
						τ ²	I ²	Explained
Without moderators		644	101			2.8%	42.9%	n.a.
U-87 MG source	cat.	644	101	0.075	n.s.	2.6%		
U-87 MG authentication	cat.	644	101	0.476	n.s.	2.8%		
U-87 MG age (cell passages)	cont.	138	11	0.238	n.s.	1.5%		
Cell concentration	cont.	113	20	0.323	n.s.	3.8%		
Confluence level at cell passaging	cont.	57	11	0.319	n.s.	0.8%		
Glucose level of culture medium	cat.	644	101	0.016	7.0%	2.5%*	40.1%	10.9%
Mycoplasma exclusion	cat.	644	101	0.491	n.s.	2.8%		
Supplemented antibiotics	cat.	644	101	0.094	n.s.	2.6%		
FBS source	cat.	644	101	0.067	n.s.	2.6%		
Type of untreated control	cat.	644	101	0.370	n.s.	2.8%		
Articles reporting quality	int.	644	101	0.031	3.3%	2.6%†	41.7%	5.0%
TMZ conc.	cont.	644	101	<0.001	38.6%	3.4%‡	64.3%	0.0%
Treatment duration	cont.	644	101	<0.001	6.0%	2.9%§	45.7%	0.0%

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 R² indicates the regression model fit⁶⁹; τ²: estimator of the variance of true effects; τ: square root of τ²; I²: proportion of between-articles variance of the total observed variance, including sampling error. τ² estimator: restricted-maximum likelihood. The column ‘Explained’ indicates the reduction of τ² after including the particular moderator compared with τ² without moderators (only applicable if the number of included effects and articles is identical). 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 τ² between parameters with different numbers of belonging articles and effects. ‘Not reported’ was included as a category for categorical moderators. P value is given for the test of the moderator. R², I² and the explained heterogeneity were only calculated for moderators that prove significance in the test of the moderator (α=0.05).

*95% CI of τ²: 1.7% to 3.6%.

†95% CI of τ²: 1.8% to 3.8%.

‡95% CI of τ²: 2.5% to 4.8%.

§95% CI of τ²: 2.1% to 4.2%.

cat., categorical; cont., continuous; FBS, fetal bovine serum; int., interval; n.a., not applicable; n.s., not significant; TMZ conc., temozolomide concentration; U-87 MG, Uppsala-87 Malignant Glioma.

effects before they could reasonably be expected to occur. This could lead to an underestimation of the effect of TMZ, or an overestimated effect of new drug candidates compared with TMZ, if the new drug candidates have an earlier onset of their effect.

Limited reporting of key statistical properties such as the number of experimental units (not reported in 24 of 137 articles) or the type of error presented in results (not reported in 17 articles) is of concern, and we note recommendations for improved reporting of such items.⁴⁶

To introduce sufficient independence between repetitions of cell culture experiments, it has been suggested that each experiment should be conducted on a different day, with freshly prepared materials, and that the experimental unit defined as the day, so n is taken as the number of days.^{47 48} Along with the limited reporting of

the number of independent experiments and (technical) replications per experiment, this leads us to encourage researchers to clearly describe their methods for introducing robust independence.

Despite the known problems with the provenance of this cell line¹⁷ and the widely recommended implementation of cell line authentication,^{49 50} we were surprised that the great majority (90.5%) of included articles did not report such an identification procedure. In addition, infrequent reporting of U-87 MG cell passage number and cell concentration used in the model adds to concern that published in vitro glioma research may be particularly confounded, as it is known that both parameters are potent drivers for heterogeneity.^{51 52} Importantly, our analysis identified an encouraging but slow trend of improving reporting over time. This is in line with recent

**Table 6** Multivariable meta-regressions

Moderators	P value	Marginal R ²	Within-articles variance		Between-articles variance	
			τ ²	Adjusted I ²	τ ²	Adjusted I ²
Without moderators			3.6%	56.6%	2.8%	42.9%
TMZ concentration and treatment duration	<0.001	42.1%	1.7% (1.5% to 1.9%)	30.9%	3.7% (2.7% to 5.2%)	68.5%
TMZ concentration, treatment duration and mediums' glucose level	<0.001	45.4%	1.7% (1.5% to 1.9%)	31.9%	3.5% (2.6% to 5.0%)	67.4%
TMZ concentration, treatment duration and articles reporting quality	<0.001	44.1%	1.7% (1.5% to 1.9%)	31.6%	3.6% (2.6% to 5.0%)	67.8%
TMZ concentration, treatment duration, mediums' glucose level and articles reporting quality	<0.001	45.9%	1.7% (1.5% to 1.9%)	32.0%	3.5% (2.5% to 5.0%)	67.4%

Multivariable 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. P value is given for the test of the moderator. Marginal R² indicates the regression model fit;⁶⁹ τ²: estimator of the variance of true effects; adjusted I²: proportion of within-articles and between-articles variance of true effects, respectively, of the total observed variance including sampling error with the indicated moderators. τ² estimator: restricted-maximum likelihood. For all multivariable meta-regressions, 644 effects in 101 articles were included. TMZ, temozolomide.

findings of a large study showing that methodological reporting quality for 1578964 PubMed Central articles had increased in recent times.⁵³ Interestingly, although showing statistical significance, the positive correlation of elevated impact factor of the journals the articles are published in and reporting quality was limited, again consistent with the findings of Menke *et al.* We note also that the anti-cell growth effect of TMZ was greater in publications which had less complete reporting of experimental details, consistent with previous work on *in vivo* studies which showed a higher effect of a therapeutic intervention reported in articles with worse methodological reporting and higher risks of bias.⁵⁴

While rigid compliance to reporting guidelines may be seen by some as unduly burdensome, our findings suggest that adoption of guidelines for the design, conduct, analysis and reporting of *in vitro* research such as the materials design analysis reporting (MDAR) framework⁴⁶ would lead to a more valuable *in vitro* research.

Sources and amplitudes of heterogeneity

The observed heterogeneity was far in excess of that expected from random sampling error ($p < 0.001$), even though we had taken steps to include broadly similar studies (outcome measurement and culture medium). As the mean effect estimate of TMZ—across all articles with all applied drug concentrations and treatment durations—is a cell viability reduction of 33.8% compared with the untreated control, the magnitude of the SD of the effects across the articles with $\pm 16.6\%$ is almost half as high as the effect estimate itself. These strongly heterogeneous findings are in line with earlier quantifications of results repeatability in cancer research^{2,8} and with the 'reproducibility project: cancer biology'.⁵⁵ These investigations evaluated reproducibility in a one-on-one replication attempt and calculated the share of reproducible

articles as the measure of reproducibility in a field. In contrast, our meta-analysis retrospectively extracted every published effect of a commonly performed experiment and calculated the variance of the effects across the articles as a measure of irreproducibility. We believe that this strategy holds an advantage in that it enables us to recognise which study parameters may act as drivers of irreproducibility. Furthermore, we think our approach is closer to scientific reality, since most studies are carried out relatively independent of each other and are not the focus of one-to-one replication of selected previously published results.

The effect of TMZ was almost 40% lower in the high glucose group than in the articles with an unreported glucose level. It is known that glucose restriction leads to a reduced cell proliferation⁵⁶ and to a sensitisation of glioblastoma cells to TMZ,^{57,58} so we consider it likely that the articles with unreported glucose levels mainly used low glucose conditions. However, as only two articles included in the meta-analysis reported the use of low glucose levels, we were not able to obtain precise enough estimates for the effect in the low glucose group. For the other experimental parameters tested in meta-regressions, no significant moderation was observed. The results do not demonstrate that there was no moderation of reproducibility and may simply reflect the statistical power of our approach depending on a minimum frequency of reporting of a particular parameter tested for its impact on reproducibility. For example, only 11 studies could be included in the meta-regression analysing the relevance of different levels of cell culture confluence at time of cell passaging, and we acknowledge this is a limitation of our study.

Implications for future research

Despite the existence of *in vitro* cell culture experimental guidelines like the Guidance on Good Cell Culture Practice,⁵² there are no widely applied reporting guidelines specific to *in vitro* preclinical research, although the MDAR framework includes *in vitro* research. Randomised group allocation, blinded outcome assessment and sample size calculation are well-established methods to reduce risks of bias in clinical and *in vivo* research, but in this review none of the included articles reported any one of these methods. This is consistent with previous findings,^{59 60} and some have argued for the implementation of randomisation and blinding in *in vitro* trials.^{2 61} Care would be required to mitigate any additional risks due to pipetting errors (because of more complex pipetting schemes) and more challenging data transfers,⁵² but the risk of unconscious systematic bias in cell plating and pipetting on multiwell plates might decrease.⁶² Although, we recommend random allocation of wells to experimental groups, blinding of cell culture procedures and assessment of cell viability to reduce potential bias. Meaningful sample size calculations will require better understanding of the experimental unit, and we endorse the suggestion that sufficient independence between replicates should be introduced by performing experiments on different days with freshly prepared cells and reagents.^{47 48 63}

The choice of control in cell culture drug response assays is important, and we were concerned that the exact condition of the untreated control arm was not reported in the majority of cases. Of note, caution is advised in the selection of maximum volume percentages of common control treatments, that is, DMSO, as elevated dosage causes inhibition of cell proliferation, exposing risk for efficacy normalisation of an investigated intervention.⁵² Where TMZ had been used as an active control for the evaluation of new therapeutic candidates, the parameters for the new drug were often much more detailed described than those for the control treatment with TMZ. As an example of why this might be important, if the effects of different drugs are differentially sensitive to glucose levels, this may lead to erroneous interpretations of the potency of a new investigational drug.

Limitations

Our study has several limitations. We did not choose the parameters contributing to the reporting quality based on a pre-existing reporting guideline because we could not identify an appropriate guideline for this type of experiment. Instead, we used parameters derived from previous work and from laboratory experience. It is unlikely that the chosen parameters have the same impact on heterogeneity, but we had no basis to assert their different impact, and so used an unweighted score.

The overwhelming part of heterogeneity of results across the articles (89.1%) remained unexplained. As discussed earlier, the main limiting factor during the analysis was the surprisingly low frequency of reporting

of experimental parameters of interest, which will have limited the power of our analysis. Further reasons for unexplained heterogeneity include contributions to irreproducibility by parameters not included in the review, or different behaviours of U-87 MG cells in different laboratories for reasons unrelated to study design.

Although our study included analysis on reporting of the source, authentication and cell passage, we are aware that a larger analysis with multiple different cancer cell models would be required to draw firmer conclusions. We think that the similar use of other glioblastoma cell lines means that our findings are probably transferable to these models.

Since the initial literature search was conducted in August 2020, we implemented an updated systematic search conducted in March 2022 and identified—besides more frequent conflicts of interest disclosures—no substantial changes in the methodological reporting habits in *in vitro* glioma research. This indicates that our findings reflect the contemporary practice in this scientific field, probably beyond the lines of neuro-oncologic research and may also provide a baseline for further projects assessing the reporting standards and reproducibility in different aspects of (*in vitro*) cancer research. Importantly, the reporting of glucose level in cell culture medium has not improved, highlighting the actuality of this review. The raise in reporting conflicts of interest may have been the result of a wider application of updated conflicts of interest disclosure guidelines by the International Committee on Medicine Journals Editors.⁶⁴

Finally, we had planned to conduct a parallel review of research using more contemporary glioblastoma *in vitro* models, such as 3D glioma stem-like cells (GSC) approaches (eg, NCH421k⁶⁵). However, in preliminary searches, we found that there was no commonly used GSC model, with authors generally using individually generated cell lines. It seemed not reasonable to perform a meta-analysis on these limited data. However, as these GSC models are probably better representatives of disease pathophysiology,^{66 67} accurate and comprehensive reporting on experimental parameters may be even more important to ensure reproducibility of their results, as elevated genetic and cellular complexity of these cells may translate into larger intrinsic biological variations.

CONCLUSION

In vitro glioma research suffers from insufficient reporting of methods and experimental design. We believe that current publication practices contribute as one source of variance that may be a driver for poor reproducibility. Although our analysis contrasts current practice with an idealised scenario and must be considered with caution in some regards (ie, risks of bias), our study clearly supports the establishment of consensus reporting guidelines for *in vitro* (cancer) research. Our review should be considered as an independent confirmatory study of earlier reporting and reproducibility enhancing recommendations⁵²



with the additional benefits of evidence for insufficient methodological reporting as well as quantification of the reproducibility of results in a highly relevant area of in vitro brain cancer research. It may be relevant for raising further awareness in a wider audience of stakeholders in biomedical research.

Acknowledgements We thank Renfei Du (Department of Neurosurgery, University Hospital Düsseldorf, Germany) for his assistance in literature screening and data extraction. Moreover, we are grateful to Shinichi Nakagawa (Department of Earth and Environmental Sciences, University of New South Wales, Sydney, Australia), Igor Fischer (Department of Neurosurgery, University Hospital Düsseldorf, Germany) and Holger Schwender (Institute of Mathematics, University of Duesseldorf, Germany) for their comments on the (meta-analytical) statistics applied. We would also like to thank Daniel Rickert (BMFZ - Genomics & Transcriptomics Laboratory, Duesseldorf, Germany) for his valuable advice on the analysis.

Contributors All authors meet the criteria for authorship as stated by the CRediT contribution for each individual author above. This includes the creation of the initial manuscript and the revision 1 and 2. TS acquired the data, conducted statistical analyses, was involved in the conceptualisation of the study and prepared the first draft of the manuscript. JG and EW assisted in data acquisition and overall methodology. SM, MM and UDK conceptualised and supervised the study. Moreover, they reviewed and edited the first draft. TS, JG and EW screened literature and extracted data for the updated systematic search. TS is the guarantor for the work and conduct of the study.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The statistical code and its datasets (initial systematic search and updated search) are publicly available on the GitHub data repository (<https://github.com/TimoSander/Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research>, License: CC-BY Attribution 4.0 International). Moreover, additional data to the manuscript (including supporting figures, tables, and detailed updated search results) are permanently publicly available at the Open Science Framework (<https://osf.io/knhx8/>, DOI: 10.17605/OSF.IO/KNHX8, License: CC-BY Attribution 4.0 International). The data are readily accessible and have no limitations for viewing. In addition, an ex-ante study protocol specifying methods and objectives of the review is also available at the Open Science Framework (<https://osf.io/u6yhb/>, DOI: 10.17605/OSF.IO/U6YHB, License: CC-BY Attribution 4.0 International).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open Practices



Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.



Open data The data and materials are available at <https://github.com/TimoSander/Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research> <https://github.com/TimoSander/Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research>.



Open material The data and materials are available at <https://github.com/TimoSander/Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research> <https://github.com/TimoSander/Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research>.



Pre-registration The systematic review and meta-analysis was formally preregistered on the Open Science Framework <https://osf.io/9k3dq>.

ORCID iDs

Timo Sander <http://orcid.org/0000-0002-5909-1184>

Joly Ghanawi <http://orcid.org/0000-0002-7945-2055>

Emma Wilson <http://orcid.org/0000-0002-8100-7508>

Sajjad Muhammad <http://orcid.org/0000-0002-6734-7979>

Malcolm Macleod <http://orcid.org/0000-0001-9187-9839>

Ulf Dietrich Kahlert <http://orcid.org/0000-0002-6021-1841>

REFERENCES

- Merton RK. The sociology of science, 1973. Available: <https://press.uchicago.edu/ucp/books/book/chicago/S/bo28451565.html> [Accessed 11 Aug 2021].
- Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012;483:531–3.
- Glenn BC, Ioannidis John PA. Reproducibility in science. *Circ Res* 2015;116:116–26.
- Freedman LP, Venugopalan G, Wisman R. Reproducibility2020: progress and priorities. *F1000Res* 2017;6:604.
- Global biological Standards Institute (GBSI). The case for standards in life science research seizing opportunities at a time of critical need, 2013. Available: https://www.academia.edu/35147125/The_Case_for_Standards_in_Life_Science_Research_Seizing_Opportunities_at_a_Time_of_Critical_Need [Accessed 9 Apr 2021].
- Hirsch C, Schildknecht S. *In Vitro* Research Reproducibility: Keeping Up High Standards. *Front Pharmacol* 2019;10:1484.
- Jarvis MF, Williams M. Irreproducibility in preclinical biomedical research: perceptions, uncertainties, and knowledge gaps. *Trends Pharmacol Sci* 2016;37:290–302.
- Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 2011;10:712.
- Wen H, Wang H-Y, He X, *et al.* On the low reproducibility of cancer studies. *Nat Sci Rev* 2018;5:619–24.
- Tamimi AF, Juweid M. Epidemiology and outcome of glioblastoma. in: de Vleeschouwer S (ED) glioblastoma. Brisbane (Au): codon publications, 2017. Available: <http://www.ncbi.nlm.nih.gov/books/NBK470003/> [Accessed 9 Apr 2021].
- Tan AC, Ashley DM, López GY, *et al.* Management of glioblastoma: state of the art and future directions. *CA Cancer J Clin* 2020;70:299–312.
- Hirst TC, Vesterinen HM, Sena ES, *et al.* Systematic review and meta-analysis of temozolomide in animal models of glioma: was clinical efficacy predicted? *Br J Cancer* 2013;108:64–71.
- Robertson FL, Marqués-Torrejón M-A, Morrison GM, *et al.* Experimental models and tools to tackle glioblastoma. *Dis Model Mech* 2019;12. doi:10.1242/dmm.040386. [Epub ahead of print: 06 09 2019].
- Poon MTC, Bruce M, Simpson JE, *et al.* Temozolomide sensitivity of malignant glioma cell lines – a systematic review assessing consistencies between in vitro studies. *BMC Cancer* 2021;21.
- Pontén J, Macintyre EH. Long term culture of normal and neoplastic human glia. *Acta Pathol Microbiol Scand* 1968;74:465–86.
- HTB-14TM, ATCC®. U-87 Mg ATCC® HTB-14TM Homo sapiens brain likely glioblastom, 2021. Available: https://www.lgcstandards-atcc.org/Products/All/HTB-14.aspx?geo_country=de [Accessed 29 Apr 2021].
- Allen M, Bjerke M, Edlund H, *et al.* Origin of the U87MG glioma cell line: good news and bad news. *Sci Transl Med* 2016;8:354re3.

- 18 Sander T, Wilson E, Ghanawi J. *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?* 2020.
- 19 Sander T. *Additional data - Meta-analysis on reporting practices as a source of heterogeneity in in vitro cancer research*, 2022.
- 20 Web of Science Group. Journal Impact Factor - Clarivate Journal Citation Reports. Web of Science Group., 2020. Available: <https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports/> [Accessed 9 Jun 2021].
- 21 Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011;342:d549.
- 22 Cheung MW-L. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychol Methods* 2014;19:211–29.
- 23 Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Res Synth Methods* 2010;1:39–65.
- 24 Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* 1977;72:320–38.
- 25 Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods* 2019;10:83–98.
- 26 Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the Random-Effects model. *J Educ Behav Stat* 2005;30:261–93.
- 27 Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;172:137–59.
- 28 Hill A. The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *J Physiol* 1910;40.
- 29 Roy Rosenzweig Center for History and New Media. Zotero [Computer software], 2021. Available: <https://www.zotero.org/>
- 30 Hair K. R. *RDedup*, 2019. Available: <https://github.com/kaitlynhair/RDedup>
- 31 Bahor Z, Liao J, Currie G, et al. Development and uptake of an online systematic review platform: the early years of the CAMARADES systematic review facility (SyRF). *BMJ Open Sci* 2021;5:e100103.
- 32 Rohatgi A. *WebPlotDigitizer*. pacifica, California, USA, 2020. Available: <https://automeris.io/WebPlotDigitizer>
- 33 RStudio Team. *RStudio: integrated development environment for R*. Boston, MA: RStudio, PBC, 2020. <http://www.rstudio.com/>
- 34 R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing, 2020. Available: <https://www.R-project.org/>
- 35 Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw* 2010;36:1–48.
- 36 Pustejovsky J. clubSandwich: Cluster-Robust (sandwich) variance estimators with Small-Sample corrections, 2021. Available: <https://CRAN.R-project.org/package=clubSandwich>
- 37 Nakagawa S, Lagisz M, O’Dea RE, et al. The orchard plot: cultivating a forest plot for use in ecology, evolution, and beyond. *Res Synth Methods* 2021;12:4–12.
- 38 Ritz C, Baty F, Streibig JC, et al. Dose-Response analysis using R. *PLoS One* 2015;10:e0146021.
- 39 Sander T. Data: Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research. Github Repository, 2022. Available: <https://github.com/TimoSander/Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research>
- 40 Jackson S, Weingart J, Nduom EK, et al. The effect of an adenosine A_{2A} agonist on intra-tumoral concentrations of temozolomide in patients with recurrent glioblastoma. *Fluids Barriers CNS* 2018;15:2.
- 41 Portnow J, Badie B, Chen M, et al. The neuropharmacokinetics of temozolomide in patients with resectable brain tumors: potential implications for the current approach to chemoradiation. *Clin Cancer Res* 2009;15:7092–8.
- 42 Ostermann S, Csajka C, Buclin T, et al. Plasma and cerebrospinal fluid population pharmacokinetics of temozolomide in malignant glioma patients. *Clin Cancer Res* 2004;10:3728–36.
- 43 Stepanenko AA, Chekhonin VP. On the critical issues in temozolomide research in glioblastoma: clinically relevant concentrations and MGMT-independent resistance. *Biomedicines* 2019;7:92.
- 44 ATCC, PBCF & PS-OC Bioresource Core Facility. *SOP: thawing, propagation and cryopreservation of NCI-PBCF-HTB14 (U-87 MG)*. 24, 2012.
- 45 Weller M, Rieger J, Grimmel C, et al. Predicting chemoresistance in human malignant glioma cells: the role of molecular genetic analyses. *Int J Cancer* 1998;79:640–4.
- 46 Macleod M, Collings AM, Graf C, et al. The MDAR (materials design analysis reporting) framework for transparent reporting in the life sciences. *Proc Natl Acad Sci U S A* 2021;118. doi:10.1073/pnas.2103238118. [Epub ahead of print: 27 04 2021].
- 47 Emmerich C. Accurate design of in vitro experiments – why does it matter? paasp network, 2016. Available: <https://paasp.net/accurate-design-of-in-vitro-experiments-why-does-it-matter/> [Accessed 24 Jun 2021].
- 48 Lazić SE, Clarke-Williams CJ, Munafò MR. What exactly is 'N' in cell culture and animal experiments? *PLoS Biol* 2018;16:e2005282.
- 49 Capes-Davis A, Bairoch A, Barrett T, et al. Cell lines as biological models: practical steps for more reliable research. *Chem Res Toxicol* 2019;32:1733–6.
- 50 International Cell Line Authentication Committee. ANSI/ATCC ASN-0002-2011 - Authentication of Human Cell Lines: Standardization of STR Profiling, 2012. Available: <https://webstore.ansi.org/standards/atcc/ansiatccasn00022011> [Accessed 16 Apr 2021].
- 51 Güllden M, Jess A, Kammann J, et al. Cytotoxic potency of H₂O₂ in cell cultures: impact of cell concentration and exposure time. *Free Radic Biol Med* 2010;49:1298–305.
- 52 OECD. Guidance document on good in vitro method practices (GIVIMP) | en | OECD, 2018. Available: <https://www.oecd.org/env/guidance-document-on-good-in-vitro-method-practices-givimp-9789264304796-en.htm> [Accessed 14 Apr 2021].
- 53 Menke J, Roelandse M, Ozyurt B, et al. The rigor and transparency index quality metric for assessing biological and medical science methods. *iScience* 2020;23:101698.
- 54 Macleod MR, van der Worp HB, Sena ES, et al. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke* 2008;39:2824–9.
- 55 eLife sciences. Reproducibility project: cancer biology. eLife, 2014. Available: <https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology> [Accessed 16 Apr 2021].
- 56 Bao Z, Chen K, Krepel S, et al. High glucose promotes human glioblastoma cell growth by increasing the expression and function of chemoattractant and growth factor receptors. *Transl Oncol* 2019;12:1155–63.
- 57 Safdie F, Brandhorst S, Wei M, et al. Fasting enhances the response of glioma to chemo- and radiotherapy. *PLoS One* 2012;7:e44603.
- 58 Wang L, Shang Z, Zhou Y, et al. Autophagy mediates glucose starvation-induced glioblastoma cell quiescence and chemoresistance through coordinating cell metabolism, cell cycle, and survival. *Cell Death Dis* 2018;9:213.
- 59 Emmerich CH, Gamboa LM, Hofmann MCJ. Improving target assessment in biomedical research: the GOT-IT recommendations. *Nat Rev Drug Discov* 2020:1–18.
- 60 NPQIP Collaborative group. Did a change in nature journals' editorial policy for life sciences research improve reporting? *BMJ Open Sci* 2019;3:e000035.
- 61 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;17:301–4.
- 62 Niepel M. A multi-center study on the reproducibility of Drug-Response assays in mammalian cell lines | Elsevier enhanced reader. Epub ahead of print 2019.
- 63 Cumming G, Fidler F, Vaux DL. Error bars in experimental biology. *J Cell Biol* 2007;177:7–11.
- 64 ICMJE. International Committee on medicine journals editors. ICMJE disclosure form., 2021. Available: <https://www.icmje.org/disclosure-of-interest/> [Accessed 3 April 2022].
- 65 Campos B, Wan F, Farhadi M, et al. Differentiation therapy exerts antitumor effects on stem-like glioma cells. *Clin Cancer Res* 2010;16:2715–28.
- 66 Lottaz C, Beier D, Meyer K, et al. Transcriptional profiles of CD133+ and CD133- glioblastoma-derived cancer stem cell lines suggest different cells of origin. *Cancer Res* 2010;70:2030–40.
- 67 Tian Y, Bresenitz P, Reska A, et al. Glioblastoma cancer stem cell lines express functional acid sensing ion channels ASIC1a and ASIC3. *Sci Rep* 2017;7:13674.
- 68 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
- 69 Nakagawa S, Schielzeth H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol Evol* 2013;4:133–42.