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

Timo Sander[1], Joly Ghanawi[2], Emma Wilson[2], Sajjad Muhammad[1], Malcolm Macleod[2] and Ulf Dietrich Kahlert[3]*

[1]Department of Neurosurgery, Heinrich-Heine University Medical Center, Düsseldorf, Germany
[2]Centre for Clinical Brain Sciences, University of Edinburgh, Scotland, United Kingdom
[3]Department of Molecular and Experimental Surgery, Clinic for General, Visceral, Vascular, and Transplant Surgery, Medical Faculty and University Hospital Magdeburg, Magdeburg, Germany

*Corresponding author mail address: mail@ulf-kahlert.com

Abstract

Background

Heterogeneity of results of exact same research experiments oppose a significant socio-economic burden. *In vitro* research presents the early step of basic science and drug development projects. 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. Glioblastoma is a form of brain cancer with largely unmet clinical need.

Methods

Here we use systematic review to describe reporting practices in *in vitro* glioblastoma research using the U87-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.

Results

In 137 identified articles, the overall methodological reporting is disappointing, e.g., the control type, mediums glucose level and cell density are reported in only 36.5, 21.2 and 16.8 percent of the articles, respectively. After adjustments for different drug concentrations and treatment durations, a three-level meta-analysis proves meaningful results heterogeneity across the studies (I^2 = 70.1%).

Conclusions

Our results further support the ongoing efforts of establishing consensus reporting practices to elevate durability of results. By doing so, we hope that this work will raise awareness of how stricter reporting may help to improve the frequency of successful translation of preclinical results into human application, not only in neuro-oncology.

Funding

We received no specific funding for this project.

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 findings either confirmed or refuted by other scientists, so that science becomes self-correcting (Merton, 1973). 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 (Begley & Ellis, 2012; Begley C. Glenn & Ioannidis John P.A., 2015; Freedman et al., 2017; Global Biological Standards Institute (GBSI), 2013; Hirsch & Schildknecht, 2019; Jarvis & Williams, 2016; Prinz et al., 2011; Wen et al., 2018).

Glioblastoma is a malignant brain tumour with a median time of survival of around 15 months (Tamimi & Juweid, 2017). First-line treatment consists of a multimodal approach of surgical resection followed by radiation therapy and chemotherapy with the alkylating agent temozolomide (TMZ) (Tan et al., 2020). 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 (Hirst et al., 2013). *In vitro* glioblastoma research commonly uses the commercially available cell line *Uppsala-87 Malignant Glioma* (U-87 MG) (Robertson et al., 2019; Poon et al., 2021), originally derived in 1966 from a 44-year-old female patient at Uppsala University (Pontén & Macintyre, 1968). However, the currently available U-87 MG line distributed by the American Type Culture Collection (ATCC, Manassas, Virginia) (HTB-14™, ATCC®, 2021) has been found to be different to the original version (Allen et al., 2016). 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

The study protocol is available at the Open Science Framework (https://osf.io/9k3dq) and was uploaded before full text based screening and data extraction began. Deviations from the protocol are described in the methods section.

Systematic Review

Systematic literature search and screening

The systematic search was conducted on the databases PubMed, Embase and Web of Science on the 26th of August 2020 using the search strategy described in Supplement 1. 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 cultures (HTB-14™, ATCC®, 2021). We only included original peer-reviewed research articles in the English language, with no restrictions made on 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. Inclusion and exclusion criteria are given in Supplement 2.

Data extraction

We recorded effect sizes for change in cell viability in response to TMZ compared to untreated control, TMZ concentration and duration of exposure. We recorded fifteen experimental parameters (Supplement 3); eight risks of bias items (Supplement 4); and the journal impact factor (JIF) for that journal in the year of publication (Web of Science Group, 2020). 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 eleven items (Supplement 5).

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 to the corresponding untreated control as raw mean difference with all data given in relation to the control viability as

$$Cell\_viability\_reduction = D = \frac{Viability_{Con} - Viability_{TMZ}}{Viability_{Con}}$$

with its variance

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

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 standard deviation (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 standard error of the mean (SEM), we assumed they were SEM, as a more conservative approach.

Multi-level random-effects meta-analysis

We used random-effects meta-analysis (Riley et al., 2011). 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 all the effects from a given article and the third level the article itself. This accounts for the relative non-independence of effects reported in the same article (Cheung, 2013). Moreover, as the exact correlations of the dependent effects within an article were unknown, we used robust-variance-estimation (Hedges et al., 2010).

To estimate tau^2, we used the restricted-maximum-likelihood method (Harville, 1977). This method has recently been shown to be robust for non-normally distributed effect sizes (Langan et al., 2019) and is recommended for the estimation of tau^2 (Viechtbauer, 2005; Langan et al., 2019). We used the t-distribution for the calculation of the weighted mean effect as this accounts for uncertainty in the estimation of tau^2 (Higgins et al., 2009). We took the within-level-three estimate of tau^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 ten parameters in univariable meta-regression which were defined a priori. For these, a reduction of within-level-three tau^2 would indicate that that parameter moderated reproducibility, with lower within-level-three tau^2 indicating that findings would be more likely to reproduce if that parameter was controlled between the original and replicating experiment. 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 (Hill, 1910). A similar four-parameter log-logistic time-response model was built 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

Software

To remove duplicate articles from the systematic search results we used two approaches, the deduplication function integrated in Zotero (Roy Rosenzweig Center for History and New Media, 2021) and one developed by the CAMARADES group (Hair, 2019). 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 (SyRF) for preclinical systematic reviews (Bahor et al., 2021). Graphically presented data were extracted with the WebPlotDigitizer (Ankit Rohatgi, 2020). Meta-analysis was performed within the RStudio environment (RStudio Team, 2020) using the programming language R (R Core Team, 2020). We used the rma.mv function of the R package Metafor (Viechtbauer, 2010) for multi-level meta-regressions, the R package clubSandwich (Pustejovsky, 2021) for robust-variance-estimation, the R package orchard (Nakagawa et al., 2021) for the calculation of marginal R^2 and I^2, the drm function of the R package drc (Ritz et al., 2015) for the dose- and time-response models and the lm function of the integrated R package stats (R Core Team, 2020) for linear regressions. The full R code and datasets are available on GitHub (https://github.com/TimoSander/Reporting-practices-as-a-source-of-heterogeneity-in-in-vitro-cancer-research/).

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). The Supplement 6 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).

Figure 1: ***Systematic search and screening results***

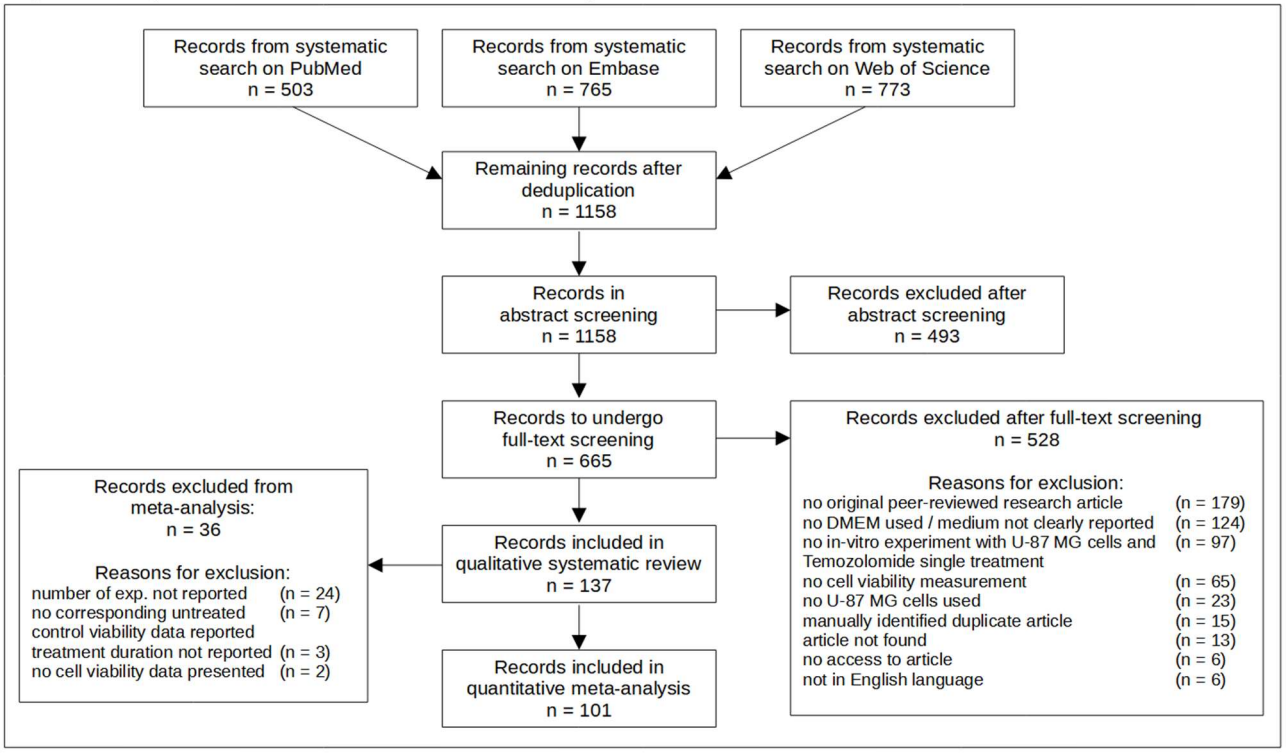


Fig. 1: Presentation based on the PRISMA statement (Page et al., 2021). Systematic searches were conducted in August 2020. Qualitative analysis included all calculations in this paper except meta-analysis and meta-regressions. One reason for exclusion per excluded article. Exp. = experiments; DMEM = Dulbecco's modified Eagle's medium; U-87 MG = Uppsala-87 Malignant Glioma.

Experimental parameters distribution

Across 137 publications, a broad range of experimental characteristics were included. The most common source of U-87 MG cells was the ATCC (66 articles, 48.2% of all included articles; Table 1a). A cell line authentication report was available in 16 articles (11.7%) and eight articles (5.8%) described testing for mycoplasma contamination. The reported cell passage number ranged from three to one hundred (median of 15), but 123 publications (89.8%) did not report it. Only 29 of 137 articles (21.2%) reported the level of culture mediums glucose, and in these high glucose supplementation (4500 mg/dl) was most prevalent (in 24 of 29). Control treatment was dimethyl sulfoxide in 37 and culture medium alone in 13 articles; in 87 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-diphenyltetrazolium (MTT) assay (67 articles, 48.9%). The concentration of U-87 MG cells used ranged from 5 to 500 cells per µl (Table 1b), with a median of 30 cells per 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, range from 50 to 90% confluence) or on time (four articles, range from 2 to 7 days), but criteria for cell passaging were not stated in 120 studies (87.6%; Table 1c).

Overall, 98 different TMZ concentrations (10nM to 16.0mM, median 100 µM; Supplement 7) and 20 different treatment durations (4 hours to 12 days, median 3 days; Supplement 8) were reported. 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.

Table 1a: ***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	American Type Culture Collection, Manassas, Virginia	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 & Streptomycin	92	67.2%
	Other antibiotics	5	3.6%
	No antibiotics supplemented	3	2.2%
	Not reported	37	27.0%
Source of fetal bovine serum (FBS)	Thermo Fisher Scientific, Waltham, Massachusetts (including Gibco, Invitrogen & Life Technologies)	51	37.2%
	Hyclone Laboratories Inc, Logan, Utah	13	9.5%
	Sigma-Aldrich, St. Louis, Missouri	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%
Cell viability assessment method	MTT assay, colorimetric	67	48.9%
	Cell Counting Kit-8 (CCK8), colorimetric	20	14.6%
	Sulforhodamine B (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%

Table 1b: ***Extracted cell concentrations***

Cell concentration [cells/µl]	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 were given	20	14.6%

Table 1c: ***Extracted cell passaging criteria***

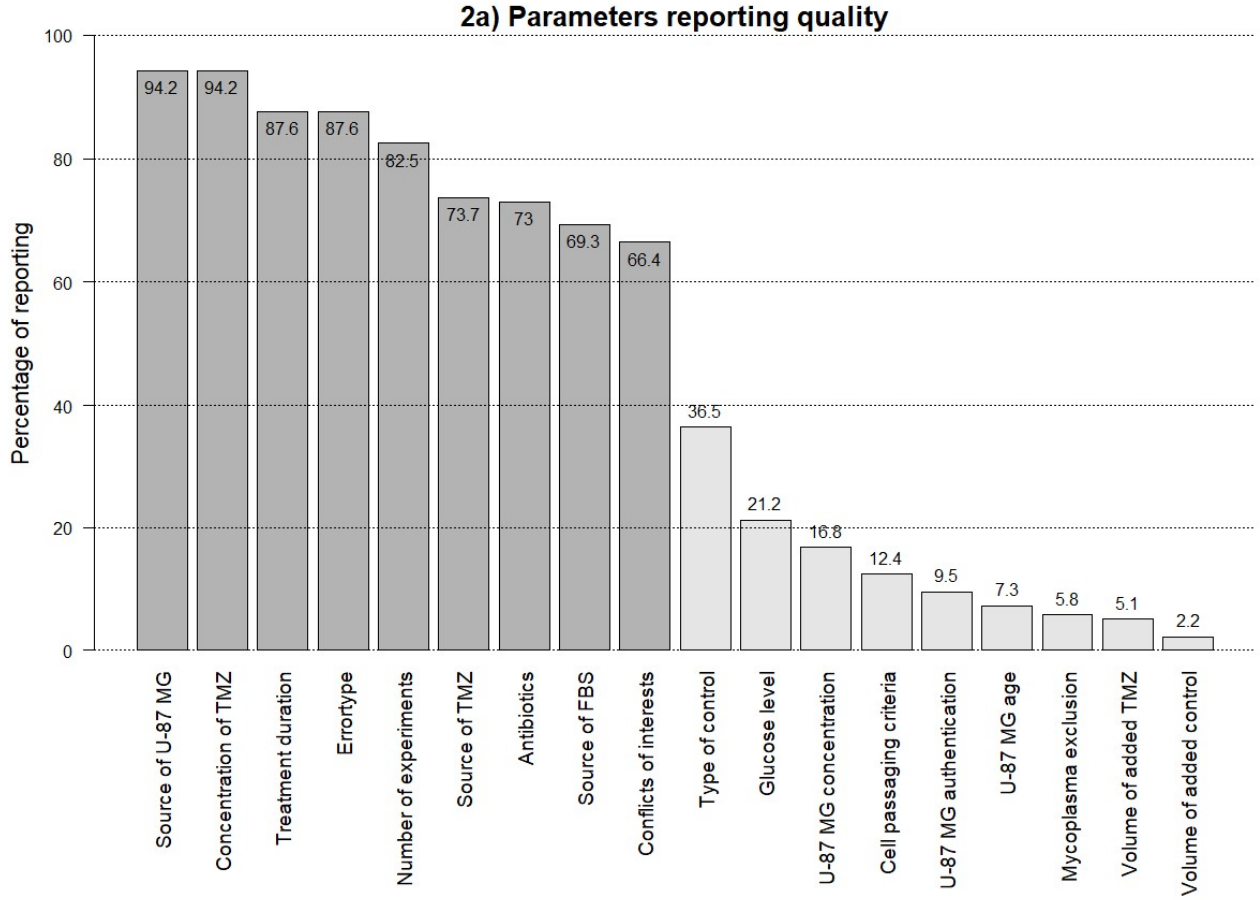
Criterion	Articles	
Based on cell culture confluence		
50% – 70%	1	0.7%
60% – 80%	1	0.7%
70%	2	1.5%
70% – 90%	3	2.2%
80%	6	4.4%
Based on time intervals		
2 days	1	0.7%
2 – 3 days	1	0.7%
3 – 4 days	1	0.7%
7 days	1	0.7%
No cell passaging criteria were reported	120	87.6%

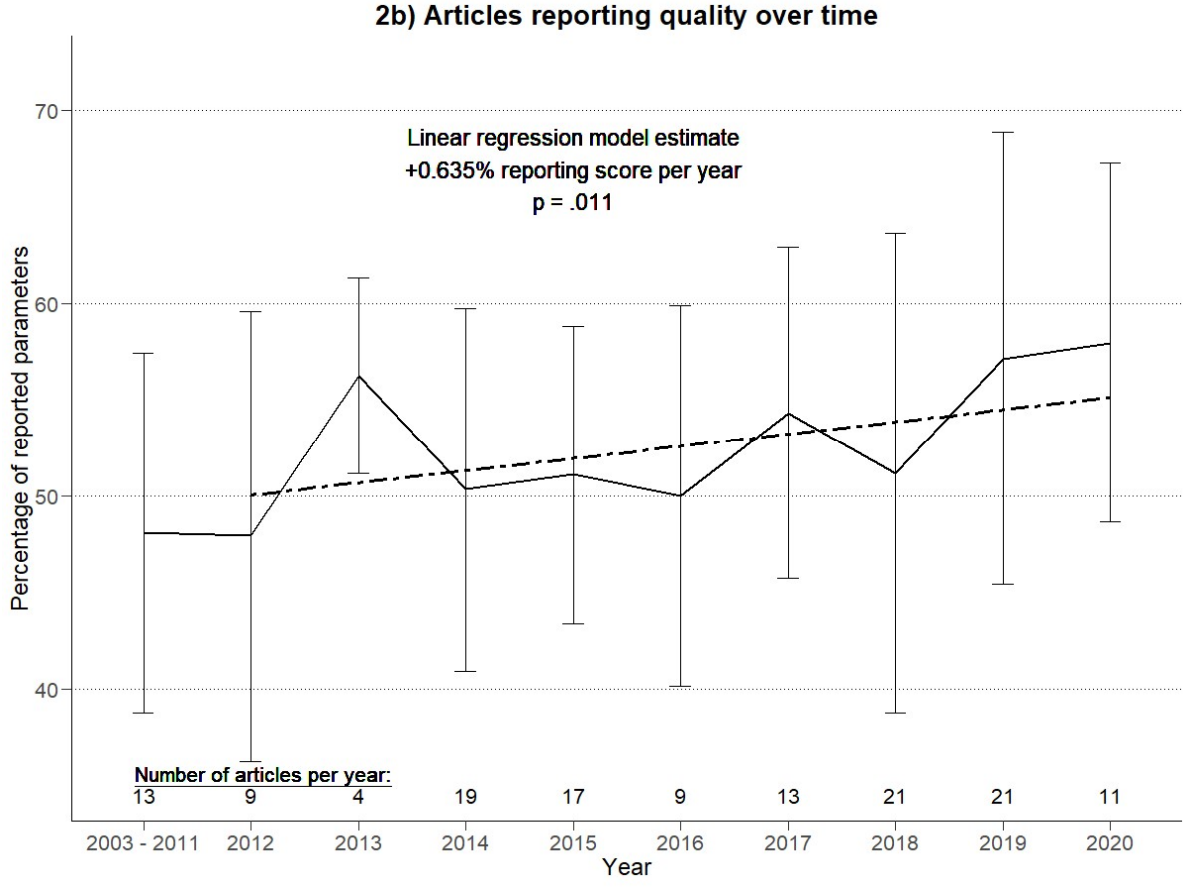
Tab. 1: Extracted parameter phenotypes (including additional information obtained through contacting the authors). One phenotype per article. The column “articles” shows the absolute and relative frequencies of articles with the parameter phenotype in relation to all 137 included articles. DMEM = Dulbecco's modified Eagle’s medium; U-87 MG = Uppsala-87 Malignant Glioma; MTT = 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide; WST-1 = Water-Soluble-Tetrazolium-1; MTS = 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium.

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 culture mediums glucose level 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 (range from three to thirteen). Analysis of change over time suggested some improvement (0.635% per year, p = .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 < .001, Figure 2c).

Figure 2: ***Reporting quality (of parameters, over time and depending on the JIF)***





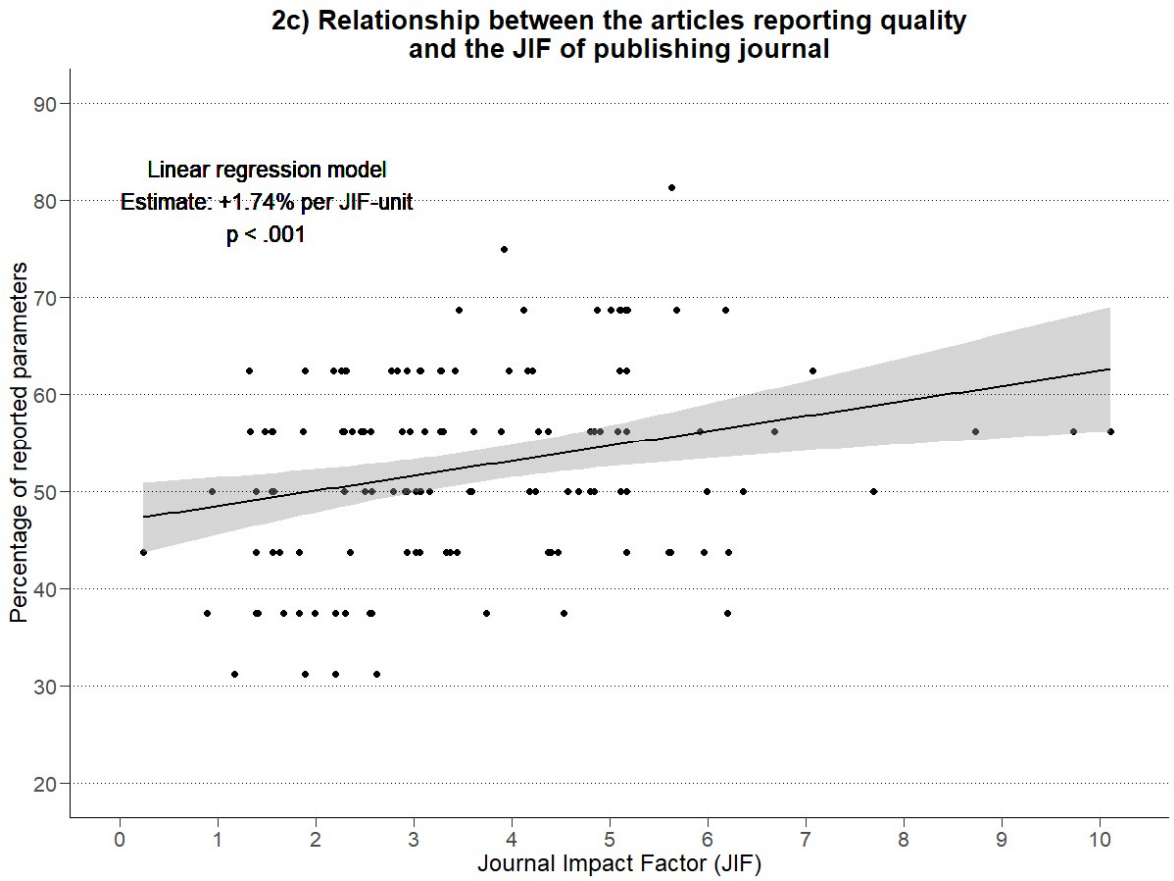


Fig. 2: a The reporting quality of a parameter was defined as the share of articles that reported this parameter phenotype in comparison to all 137 included articles. This share is shown on the top of each bar for each parameter. A parameter was considered to be reported if its phenotype was clear based on the information provided in the original full-text research article. **b** Linear regression model of articles reporting quality (proportion of reported parameters in 16 selected parameters) and year of publication. The articles published before 2012 were summarized because of low numbers of articles published in these years (but the exact years were used for the regression). Only articles until the time of systematic search in August 2020 were considered. The dotted line represents the linear regression line; error bars indicate the standard deviation 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 % confidence interval of the regression model prediction. JIF 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 time of analysis. One article was omitted as no JIF could be obtained. U-87 MG = Uppsala 87 Malignant Glioma; FBS = Fetal bovine serum; JIF = Journal impact factor; TMZ = Temozolomide.

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 pre-registered protocol specifying the hypotheses and outcomes (Table 2). 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 was unclear in 47 articles. The mean number of measures to reduce risks of bias reported was 2.9.

253 **Table 2:** ***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%

254

255 **Tab. 2:** The column "articles" shows the absolute and relative prevalence of articles having a
256 particular risk of bias factor in comparison to all 137 included articles.

257

258 # Meta-analysis of the effect of TMZ

259 The observed effect of TMZ is highly heterogeneous; variation within different experiments in the
260 same article (represented by I^2 of level-two variance) accounts for 56.6% of observed variance,
261 variation of the effect across different articles (represented by I^2 of level-three variance) accounts for
262 42.9% of observed variance, and the variance due to random chance expected if all experiments of
263 all articles were held under identical conditions made up only 0.5% of the total variance (Table 3).
264 The heterogeneity of results across the articles is reflected in a SD of +/- 16.6% (95% CI for this SD
265 estimate from 13.9% to 19.8%) around a global estimate of a reduction in cell viability following TMZ
266 treatment of 33.8% (95% CI from 30.0 to 37.7%) compared to the untreated control.

Table 3: ***Random-effects three-level meta-analysis suggests significant irreproducibility***

Included 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 an untreated control)	33.8% [30.0%, 37.7%]
Investigation of heterogeneity	
Test whether heterogeneity is present	
Q (df = 643):	134066.1
p value	< .001
Total I^2:	99.5%
Within-articles-variance of the true effect	
tau^2	3.6% [3.2%,4.1%]
tau	19.0% [17.9%,20.3%]
I^2 (= proportion of total variance)	56.6%
Between-articles-variance of the true effect (~Irreproducibility)	
tau^2	2.8% [1.9%,3.9%]
tau	16.6% [13.9%,19.8%]
I^2 (= proportion of total variance)	42.9%

Tab. 3: 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. tau^2: estimator of the variance of true effects (level-two variance = within-articles-variance; level-three variance = between-articles-variance = representant of irreproducibility); tau = square root of tau^2; I^2: proportion of within- and between-articles-variance, respectively, of the total observed variance including sampling error. tau^2 estimator: restricted-maximum likelihood. Cochran's Q was used as the test for heterogeneity using a chi-squared distribution; values in square brackets show confidence intervals with significance level set at 0.05. df = degrees of freedom; TMZ = Temozolomide; U-87 MG = Uppsala-87 Malignant Glioma.

Drivers of heterogeneity

The within-articles-variance of effects reported in the same article could be, as expected, partly explained by differences in TMZ concentrations (50.7%) and treatment durations (5.0%) (Table 4). 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 tau^2 from 3.6% to 1.6% while the estimated between-articles-variance increased from 2.8% to 3.8% (Table 6).

The glucose level applied in the cell culture medium was the only experimental parameter significantly associated with heterogeneity of results across the articles (p = .016, Table 5). The moderator fit indicated by marginal R^2 was 7.0%, and 10.9% of the between-articles-variance tau^2 could be explained because of differences in the glucose level. In other words, roughly 11 percent of heterogeneous results were attributed to different glucose levels used in the culture medium. The different glucose levels and effects are shown in Table 7, indicating a significantly smaller effect of TMZ in the articles using high glucose supplementation (cell viability reduction of 23.1% compared

293 to the untreated control, 95% CI from 15.8 to 30.5%) than in the articles using an unreported glucose
294 level (37.1%, 95% CI from 28.6 to 45.6%, p = .002).

295 Articles reporting quality showed a significant correlation with the reported effect of TMZ (marginal
296 R^2 = 3.3%). Reported effect on cell viability fell by 3.0% for each unit increase in the number of
297 reported parameters (p = .026). Adding these two features (glucose level and articles reporting
298 quality) to the multivariable between-articles-variance meta-regression resulted in a slightly
299 improved model, reducing tau^2 from 3.8% to 3.6%.

Table 4: ***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	39.9%	1.8%[c]	33.1%	50.7%
Treatment duration	cont.	644	101	< .001	6.4%	3.4%[d]	53.7%	5.0%

Tab. 4: 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 (Nakagawa & Schielzeth, 2013); 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%.2.1%]; **d**: 95%-CI of tau^2: [3.0%,3.9%].

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 tau^2	I^2	Explained
Without moderators		644	101			2.8%	42.9%	n. a.
U-87 MG source	cat.	644	101	.075	n. s.	2.6%		
U-87 MG authentication	cat.	644	101	.476	n. s.	2.8%		
U87-MG age (Cell passages)	cont.	138	11	.238	n. s.	1.5%		
Cell concentration	cont.	113	20	.323	n. s.	3.8%		
Confluence level at cell passaging	cont.	57	11	.319	n. s.	0.8%		
Glucose level of culture medium	cat.	644	101	.016	7.0%	2.5%[a]	40.1%	10.9%
Mycoplasma exclusion	cat.	644	101	.491	n. s.	2.8%		
Supplemented antibiotics	cat.	644	101	.094	n. s.	2.6%		
FBS source	cat.	644	101	.067	n. s.	2.6%		
Type of untreated control	cat.	644	101	.370	n. s.	2.8%		
Articles reporting quality	int.	644	101	.031	3.3%	2.6%[b]	41.7%	5.0%
TMZ conc.	cont.	644	101	< .001	39.9%	3.6%[c]	66.3%	0.0%
Treatment duration	cont.	644	101	< .001	6.4%	2.9%[d]	45.7%	0.0%

Tab. 5: 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 (Nakagawa & Schielzeth, 2013); tau^2: estimator of the variance of true effects; tau = square root of tau^2; I^2: proportion of between-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 and I^2 and the explained heterogeneity were only calculated for moderators that prove significance in the test of the moderator (alpha = .05). n. s. = not significant. TMZ conc. = Temozolomide concentration. **a**: 95%-confidence-interval (CI): tau^2: [1.7%,3.6%]; **b**: 95%-CI of tau^2: [1.8%,3.8%]; **c**: 95%-CI of tau^2: [2.6%,5.0%]; **d**: 95%-CI of tau^2: [2.1%,4.2%]

Table 6: ***Multivariable meta-regressions***

Moderators	*p* value	Marg. R^2	Within-articles-variance tau²	Within-articles-variance adjusted I^2	Between-articles-variance tau²	Between-articles-variance adjusted I^2
Without moderators			3.6%	56.6%	2.8%	42.9%
TMZ concentration & treatment duration	< .001	43.5%	1.6% [1.4%,1.8%]	29.3%	3.8% [2.8%,5.3%]	70.1%
TMZ concentration & Treatment duration & mediums glucose level	< .001	47.0%	1.6% [1.4%,1.8%]	30.4%	3.6% [2.6%,5.0%]	68.9%
TMZ concentration & Treatment duration & articles reporting quality	< .001	45.6%	1.6% [1.4%,1.8%]	30.0%	3.7% [2.7%,5.1%]	69.4%
TMZ concentration & Treatment duration & mediums glucose level & articles reporting quality	< .001	47.5%	1.6% [1.4%,1.8%]	30.5%	3.6% [2.6%,5.0%]	68.9%

Tab. 6: 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. The *p* value is for the test of the moderator. Marginal R^2 indicates the regression model fit (Nakagawa & Schielzeth, 2013); tau²: estimator of the variance of true effects; adjusted I^2: proportion of within- and between-articles-variance of true effects, respectively, of the total observed variance including sampling error with the indicated moderators. tau² estimator: restricted-maximum likelihood. For all multivariable meta-regressions, 644 effects in 101 articles were included. TMZ = Temozolomide.

Table 7: ***Correlation between the effect and culture mediums glucose level as well as articles reporting quality***

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

Tab. 7: 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. SD = standard deviation; CI = confidence interval (with significance level of alpha = .05); LL = lower limit; UL = upper limit. *: 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.

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 ($p < .001$). 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 (Poon et al., 2021).

Based on TMZ concentrations found in peritumoral tissue (Jackson et al., 2018; Portnow et al., 2009) and cerebrospinal fluid (Ostermann et al., 2004), Stepanenko and Checkhonin recommended *in vitro* studies should use concentrations of 1 to 10 µM (Stepanenko & Chekhonin, 2019). Although the effect of higher concentrations could be used, 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 internalization, which usually occurs during cell division. An effect of TMZ can therefore be expected at the earliest after 1.5 cell doubling times, and the cell doubling time of U-87 MG is around 34 hours (ATCC, PBCF & PS-OC Bioresource Core Facility, 2012; Weller et al., 1998). Applying an early credible limit for efficacy of 51 hours, 31.7% of the articles included in the meta-analysis only measured 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 to TMZ, if the new drug candidates have an earlier onset of effect.

Limited reporting of key statistical properties such as the number of experimental units included (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 (Macleod et al., 2021).

To introduce sufficient independence between repetitions of cell culture experiments, it has been suggested that experiments should be conducted over several days, with freshly prepared materials, and that the experimental unit defined as the day, so n is taken as the number of days (C. Emmerich, 2016; Lazic et al., 2018). 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 line (Allen et al., 2016) and the widely recommended implementation of cell line authentication (Capes-Davis et al., 2019; International Cell Line Authentication Committee, 2012), 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 (Gülden et al., 2010; OECD, 2018). Importantly, our analysis identified an encouraging but slow trend of improving reporting over time. This is in line with recent findings of a large study showing that methodological reporting quality for 1,578,964 PubMedCentral articles had increased in recent times (Menke et al., 2020). 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 (Macleod et al., 2008).

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 MDAR framework (Macleod et al., 2021) would lead to more useful *in vitro* research.

Sources and amplitudes of heterogeneity

The observed heterogeneity was far in excess of that expected from random sampling error ($p < .001$), even though we had taken steps to include broadly similar studies (outcome measure, 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 to the untreated control, the magnitude of the SD of the effects across the articles with +/- 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 (Begley & Ellis, 2012; Prinz et al., 2011) and with the "Reproducibility Project: Cancer Biology" (eLife sciences, 2014). 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 has an advantage in that it enables us to recognise which study parameters may act as drivers of irreproducibility. Further, we think our approach has broader relevance, 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 forty percent 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 proliferation (Bao et al., 2019) and to a sensitisation of glioblastoma cells to TMZ (Safdie et al., 2012; Wang et al., 2018), so we consider it likely that the articles with unreported glucose levels mainly used low

glucose levels. 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 effect on 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 in a meta-regression. For example, only eleven 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 based experimental guidelines like the Guidance on Good Cell Culture Practice (OECD, 2018) there are no widely applied reporting guidelines specific to *in vitro* preclinical research, although the MDAR framework includes *in vitro* research. Randomised group allocation in experimental design, 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 (C. H. Emmerich et al., 2020; The NPQIP Collaborative group, 2019), and some have argued for the implementation of randomization and blinding in *in vitro* trials (Begley & Ellis, 2012; Krithikadatta et al., 2014). Care would be required to mitigate any additional risks due to pipetting errors (because of complex pipetting schemes) and more challenging data transfers (OECD, 2018), but the risk of unconscious systematic bias in cell plating and pipetting on multi-well plates might decrease (Niepel, 2019). Although, we recommend random allocation of wells to experimental groups and the 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 (Cumming et al., 2007; C. Emmerich, 2016; Lazic et al., 2018).

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, i.e., dimethyl sulfoxide, as elevated dosage causes inhibition of cell proliferation, exposing risk for efficacy normalization of an investigated intervention (OECD, 2018). 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.

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 approaches (e.g. NCH421k (Campos et al., 2010)). However, in preliminary searches we were surprised to find 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 (Lottaz et al., 2010; Tian et al., 2017), 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 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 (i.e., risks of bias), our study clearly supports the establishment of consensus reporting guidelines for *in vitro* (cancer) research. Our study should be considered as an independent confirmatory study of earlier reporting and reproducibility enhancing recommendations (OECD, 2018) 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 to raise further awareness in a wide audience of stakeholders in biomedical research.

Acknowledgements

We thank Renfei Du (Department of Neurosurgery, University of Duesseldorf, Germany) for his assistance in literature screening and data extraction. Moreover, we are grateful to Shinichi Nakagawa (Earth and Environmental Sciences, University of New South Wales, Sydney, Australia), Igor Fischer (Department of Neurosurgery, University of Duesseldorf, Germany) and Holger Schwender (Institute of Mathematics, University of Duesseldorf, Germany) for their comments on the (meta-analytical) statistics applied.

We declare no competing interests.

Abbreviations

ATCC	American Type Culture Collection, Manassas, Virginia
CI	Confidence interval
DMEM	Dulbecco's Modified Eagle Medium
JIF	Journal impact factor
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
SD	Standard deviation
SEM	Standard error of the mean
SyRF	Systematic Review Facility
TMZ	Temozolomide
U-87 MG	Uppsala-87 Malignant Glioma

524 # References

Allen, M., Bjerke, M., Edlund, H., Nelander, S., & Westermark, B. (2016). Origin of the U87MG glioma cell line: Good news and bad news. *Science Translational Medicine*, *8*(354), 354re3-354re3. https://doi.org/10.1126/scitranslmed.aaf6853

Ankit Rohatgi. (2020). *WebPlotDigitizer* (4.4) [Computer software]. https://automeris.io/WebPlotDigitizer

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

Bahor, Z., Liao, J., Currie, G., Ayder, C., Macleod, M., McCann, S. K., Bannach-Brown, A., Wever, K., Soliman, N., Wang, Q., Doran-Constant, L., Young, L., Sena, E. S., & Sena, C. (2021). Development and uptake of an online systematic review platform: The early years of the CAMARADES Systematic Review Facility (SyRF). *BMJ Open Science*, *5*(1), e100103. https://doi.org/10.1136/bmjos-2020-100103

Bao, Z., Chen, K., Krepel, S., Tang, P., Gong, W., Zhang, M., Liang, W., Trivett, A., Zhou, M., & Wang, J. M. (2019). High Glucose Promotes Human Glioblastoma Cell Growth by Increasing the Expression and Function of Chemoattractant and Growth Factor Receptors. *Translational Oncology*, *12*(9), 1155–1163. https://doi.org/10.1016/j.tranon.2019.04.016

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. https://doi.org/10.1038/483531a

Begley C. Glenn & Ioannidis John P.A. (2015). Reproducibility in Science. *Circulation Research*, *116*(1), 116–126. https://doi.org/10.1161/CIRCRESAHA.114.303819

Campos, B., Wan, F., Farhadi, M., Ernst, A., Zeppernick, F., Tagscherer, K. E., Ahmadi, R., Lohr, J., Dictus, C., Gdynia, G., Combs, S. E., Goidts, V., Helmke, B. M., Eckstein, V., Roth, W., Beckhove, P., Lichter, P., Unterberg, A., Radlwimmer, B., & Herold-Mende, C. (2010). Differentiation Therapy Exerts Antitumor Effects on Stem-like Glioma Cells. *Clinical Cancer Research*, *16*(10), 2715–2728.

Capes-Davis, A., Bairoch, A., Barrett, T., Burnett, E. C., Dirks, W. G., Hall, E. M., Healy, L., Kniss, D. A., Korch, C., Liu, Y., Neve, R. M., Nims, R. W., Parodi, B., Schweppe, R. E., Storts, D. R., & Tian, F. (2019). Cell Lines as Biological Models: Practical Steps for More Reliable Research. *Chemical Research in Toxicology*, *32*(9), 1733–1736. https://doi.org/10.1021/acs.chemrestox.9b00215

Cheung, M. (2013). Modeling Dependent Effect Sizes With Three-Level Meta-Analyses: A Structural Equation Modeling Approach. *Psychological Methods*, *19*. https://doi.org/10.1037/a0032968

Cumming, G., Fidler, F., & Vaux, D. L. (2007). Error bars in experimental biology. *The Journal of Cell Biology*, *177*(1), 7–11. https://doi.org/10.1083/jcb.200611141

eLife sciences. (2014, December 10). *Reproducibility Project: Cancer Biology*. ELife; eLife Sciences Publications Limited. https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology

Emmerich, C. (2016, August 24). Accurate design of in vitro experiments – why does it matter? *Paasp Network*. https://paasp.net/accurate-design-of-in-vitro-experiments-why-does-it-matter/

Emmerich, C. H., Gamboa, L. M., Hofmann, M. C. J., Bonin-Andresen, M., Arbach, O., Schendel, P., Gerlach, B., Hempel, K., Bespalov, A., Dirnagl, U., & Parnham, M. J. (2020). Improving target assessment in biomedical research: The GOT-IT recommendations. *Nature Reviews. Drug Discovery*, 1–18. https://doi.org/10.1038/s41573-020-0087-3

Freedman, L. P., Venugopalan, G., & Wisman, R. (2017). Reproducibility2020: Progress and priorities. *F1000Research*, *6*, 604. https://doi.org/10.12688/f1000research.11334.1

Global Biological Standards Institute (GBSI). (2013). *The Case for Standards in Life Science Research Seizing Opportunities at a Time of Critical Need*. https://www.academia.edu/35147125/The_Case_for_Standards_in_Life_Science_Research_Seizing_Opportunities_at_a_Time_of_Critical_Need

Gülden, M., Jess, A., Kammann, J., Maser, E., & Seibert, H. (2010). Cytotoxic potency of H2O2 in cell cultures: Impact of cell concentration and exposure time. *Free Radical Biology and Medicine*, *49*(8), 1298–1305. https://doi.org/10.1016/j.freeradbiomed.2010.07.015

Hair, K. (2019). *RDedup*. https://github.com/kaitlynhair/RDedup

Harville, D. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, *72*(358), 320–338. https://doi.org/10.2307/2286796

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, *172*(1), 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x

Hirsch, C., & Schildknecht, S. (2019). In Vitro Research Reproducibility: Keeping Up High Standards. *Frontiers in Pharmacology*, *10*. https://doi.org/10.3389/fphar.2019.01484

Hirst, T. C., Vesterinen, H. M., Sena, E. S., Egan, K. J., Macleod, M. R., & Whittle, I. R. (2013). Systematic review and meta-analysis of temozolomide in animal models of glioma: Was clinical efficacy predicted? *British Journal of Cancer*, *108*(1), 64–71. https://doi.org/10.1038/bjc.2012.504

HTB-14^{TM}, ATCC®. (2021). *U-87 MG ATCC ® HTB-14^{TM} Homo sapiens brain Likely glioblastom*. https://www.lgcstandards-atcc.org/Products/All/HTB-14.aspx?geo_country=de

International Cell Line Authentication Committee. (2012). *ANSI/ATCC ASN-0002-2011—Authentication of Human Cell Lines: Standardization of STR Profiling*. https://webstore.ansi.org/standards/atcc/ansiatccasn00022011

Jarvis, M. F., & Williams, M. (2016). Irreproducibility in Preclinical Biomedical Research: Perceptions, Uncertainties, and Knowledge Gaps. *Trends in Pharmacological Sciences*, *37*(4), 290–302. https://doi.org/10.1016/j.tips.2015.12.001

Krithikadatta, J., Gopikrishna, V., & Datta, M. (2014). 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. *Journal of Conservative Dentistry : JCD*, *17*(4), 301–304. https://doi.org/10.4103/0972-0707.136338

Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W., & Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods*, *10*(1), 83–98. https://doi.org/10.1002/jrsm.1316

Lazic, S. E., Clarke-Williams, C. J., & Munafò, M. R. (2018). What exactly is 'N' in cell culture and animal experiments? *PLOS Biology*, *16*(4), e2005282. https://doi.org/10.1371/journal.pbio.2005282

Lottaz, C., Beier, D., Meyer, K., Kumar, P., Hermann, A., Schwarz, J., Junker, M., Oefner, P. J., Bogdahn, U., Wischhusen, J., Spang, R., Storch, A., & Beier, C. P. (2010). Transcriptional Profiles of CD133 and CD133− Glioblastoma-Derived Cancer Stem Cell Lines Suggest Different Cells of Origin. *Cancer Research*, *70*(5), 2030–2040.

Macleod, Collings, A. M., Graf, C., Kiermer, V., Mellor, D., Swaminathan, S., Sweet, D., & Vinson, V. (2021). The MDAR (Materials Design Analysis Reporting) Framework for transparent reporting in the life sciences. *Proceedings of the National Academy of Sciences*, *118*(17). https://doi.org/10.1073/pnas.2103238118

Macleod, M. R., van der Worp, H. B., Sena, E. S., Howells, D. W., Dirnagl, U., & Donnan, G. A. (2008). Evidence for the Efficacy of NXY-059 in Experimental Focal Cerebral Ischaemia Is Confounded by Study Quality. *Stroke*, *39*(10), 2824–2829. https://doi.org/10.1161/STROKEAHA.108.515957

Menke, J., Roelandse, M., Ozyurt, B., Martone, M., & Bandrowski, A. (2020). The Rigor and Transparency Index Quality Metric for Assessing Biological and Medical Science Methods. *IScience*, *23*(11), 101698. https://doi.org/10.1016/j.isci.2020.101698

Merton, R. K. (1973). *The Sociology of Science*. https://press.uchicago.edu/ucp/books/book/chicago/S/bo28451565.html

Nakagawa, S., Lagisz, M., O'Dea, R. E., Rutkowska, J., Yang, Y., Noble, D. W. A., & Senior, A. M. (2021). *The orchard plot: Cultivating a forest plot for use in ecology, evolution, and beyond*. https://doi.org/10.1002/jrsm.1424

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Niepel, M. (2019). *A Multi-center Study on the Reproducibility of Drug-Response Assays in Mammalian Cell Lines | Elsevier Enhanced Reader*. https://doi.org/10.1016/j.cels.2019.06.005

OECD. (2018). *Guidance Document on Good In Vitro Method Practices (GIVIMP) | en | OECD*. https://www.oecd.org/env/guidance-document-on-good-in-vitro-method-practices-givimp-9789264304796-en.htm

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pontén, J., & Macintyre, E. H. (1968). Long Term Culture of Normal and Neoplastic Human Glia. *Acta Pathologica Microbiologica Scandinavica*, *74*(4), 465–486. https://doi.org/10.1111/j.1699-0463.1968.tb03502.x

Poon, M. T., Bruce, M., Simpson, J. E., Hannan, C. J., & Brennan, P. M. (2021). *Temozolomide sensitivity of malignant glioma cell lines – a systematic review assessing consistencies between in vitro studies* (p. 2021.06.29.21259733). https://doi.org/10.1101/2021.06.29.21259733

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, *10*(9), 712–712. https://doi.org/10.1038/nrd3439-c1

Pustejovsky, J. (2021). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections* (0.5.3) [Computer software]. https://CRAN.R-project.org/package=clubSandwich

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ*, *342*, d549. https://doi.org/10.1136/bmj.d549

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-Response Analysis Using R. *PLOS ONE*, *10*(12), e0146021. https://doi.org/10.1371/journal.pone.0146021

Robertson, F. L., Marqués-Torrejón, M.-A., Morrison, G. M., & Pollard, S. M. (2019). Experimental models and tools to tackle glioblastoma. *Disease Models & Mechanisms*, *12*(dmm040386). https://doi.org/10.1242/dmm.040386

Roy Rosenzweig Center for History and New Media. (2021). *Zotero [Computer software]* (5.0.96.2) [Computer software]. https://www.zotero.org/

RStudio Team. (2020). *RStudio: Integrated Development Environment for R* (1.3.1093) [Computer software]. RStudio, PBC. http://www.rstudio.com/

Safdie, F., Brandhorst, S., Wei, M., Wang, W., Lee, C., Hwang, S., Conti, P. S., Chen, T. C., & Longo, V. D. (2012). Fasting Enhances the Response of Glioma to Chemo- and Radiotherapy. *PLOS ONE*, *7*(9), e44603. https://doi.org/10.1371/journal.pone.0044603

Tamimi, A. F., & Juweid, M. (2017). Epidemiology and Outcome of Glioblastoma. In S. De Vleeschouwer (Ed.), *Glioblastoma*. Codon Publications. http://www.ncbi.nlm.nih.gov/books/NBK470003/

Tan, A. C., Ashley, D. M., López, G. Y., Malinzak, M., Friedman, H. S., & Khasraw, M. (2020). Management of glioblastoma: State of the art and future directions. *CA: A Cancer Journal for Clinicians*, *70*(4), 299–312. https://doi.org/10.3322/caac.21613

The NPQIP Collaborative group. (2019). Did a change in Nature journals' editorial policy for life sciences research improve reporting? *BMJ Open Science*, *3*(1), e000035. https://doi.org/10.1136/bmjos-2017-000035

Tian, Y., Bresenitz, P., Reska, A., El Moussaoui, L., Beier, C. P., & Gründer, S. (2017). Glioblastoma cancer stem cell lines express functional acid sensing ion channels ASIC1a and ASIC3. *Scientific Reports*, *7*(1), 13674. https://doi.org/10.1038/s41598-017-13666-9

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261–293. https://doi.org/10.3102/10769986030003261

Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, *36*(1), 1–48. https://doi.org/10.18637/jss.v036.i03

Wang, L., Shang, Z., Zhou, Y., Hu, X., Chen, Y., Fan, Y., Wei, X., Wu, L., Liang, Q., Zhang, J., & Gao, Z. (2018). Autophagy mediates glucose starvation-induced glioblastoma cell quiescence and chemoresistance through coordinating cell metabolism, cell cycle, and survival. *Cell Death & Disease*, *9*(2), 213. https://doi.org/10.1038/s41419-017-0242-x

Web of Science Group. (2020, and previous years down to 2003). Journal Impact Factor—Clarivate Journal Citation Reports. *Web of Science Group*. https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports/

Weller, M., Rieger, J., Grimmel, C., Van Meir, E. G., De Tribolet, N., Krajewski, S., Reed, J. C., von Deimling, A., & Dichgans, J. (1998). Predicting chemoresistance in human malignant glioma cells: The role of molecular genetic analyses. *International Journal of Cancer*, *79*(6), 640–644. https://doi.org/10.1002/(sici)1097-0215(19981218)79:6<640::aid-ijc15>3.0.co;2-z

Wen, H., Wang, H.-Y., He, X., & Wu, C.-I. (2018). On the low reproducibility of cancer studies. *National Science Review*, *5*(5), 619–624. https://doi.org/10.1093/nsr/nwy021