RESEARCH ARTICLE

'Invisible actors'—How poor methodology reporting compromises mouse models of oncology: A cross-sectional survey

Elizabeth A. Nunamaker[1¤], **Penny S. Reynolds**[2]*

1 Animal Care Services, University of Florida, Gainesville, Florida, United States of America, **2** Department of Anesthesiology, Statistics in Anesthesiology Research (STAR) Core, College of Medicine, University of Florida, Gainesville, Florida, United States of America

¤ Current address: Global Animal Welfare and Training, Charles River Laboratories, Wilmington, Massachusetts, United States of America

* PReynolds@anest.ufl.edu

Abstract

The laboratory mouse is a key player in preclinical oncology research. However, emphasis of techniques reporting at the expense of critical animal-related detail compromises research integrity, animal welfare, and, ultimately, the translation potential of mouse-based oncology models. To evaluate current reporting practices, we performed a cross-sectional survey of 400 preclinical oncology studies using mouse solid-tumour models. Articles published in 2020 were selected from 20 journals that specifically endorsed the ARRIVE (Animal Research: Reporting of In Vivo Experiments) preclinical reporting guidelines. We assessed reporting compliance for 22 items in five domains: ethical oversight assurance, animal signalment, husbandry, welfare, and euthanasia. Data were analysed using hierarchical generalised random-intercept models, clustered on journal. Overall, reporting of animal-related items was poor. Median compliance over all categories was 23%. There was little or no association between extent of reporting compliance and journal or journal impact factor. Age, sex, and source were reported most frequently, but verifiable strain information was reported for <10% of studies. Animal husbandry, housing environment, and welfare items were reported by <5% of studies. Fewer than one in four studies reported analgesia use, humane endpoints, or an identifiable method of euthanasia. Of concern was the poor documentation of ethical oversight information. Fewer than one in four provided verifiable approval information, and almost one in ten reported no information, or information that was demonstrably false. Mice are the "invisible actors" in preclinical oncology research. In spite of widespread endorsement of reporting guidelines, adherence to reporting guidelines on the part of authors is poor and journals fail to enforce guideline reporting standards. In particular, the inadequate reporting of key animal-related items severely restricts the utility and translation potential of mouse models, and results in research waste. Both investigators and journals have the ethical responsibility to ensure animals are not wasted in uninformative research.

Introduction

The laboratory mouse is a well-established and common research model used to study human diseases, and a key link in the translation of bench experiments to clinical trials. In cancer research, mouse models are major players in three domains. First, mouse models enable insight into the genetic, mechanistic, and phenotypic mechanisms and interactions underlying the pathogenesis of cancer initiation and tumour formation. Second, they serve as *in vivo* platforms for drug discovery and therapeutic screening and evaluation. Finally, mouse models permit direct testing of the relationship of tumorigenesis to various environmental factors not possible in clinical studies of humans [1].

However, the 'mouse model' in oncology research is not a monolith. Mouse strains are varied, animal genotypes are manipulated, induction methods are wide-ranging (e.g. engraftment, syngeneic, orthotopic, and genetically-engineered models), and disparate methods of determining marker expression are used [1, 2]. While lab bench methods are usually well described and thoroughly documented in literature reports, the animals themselves have been 'invisible actors'. Information on routine care, housing, welfare measures (such as anaesthesia, analgesia, euthanasia, [3]), and animal signalment (strain, sub-strain, age and sex) have all been documented to influence both progression of specific cancers and expression of experimental outcomes [4]. An additional consideration is that many cancer models are associated with high rates of lethality and potential for suffering, so reporting of care and welfare measures are necessary for assessing if studies do in fact meet basic ethical standards. However, this information has not been prioritised in much of the literature. The omission of animal-related details, intentional or not, may be due in part to the perception of mice as disposable, inter-changeable commodities, or "furry test tubes" [5]. Without complete and accurate description of all methods related to the specific animal model, including care and welfare, it will not be possible to assess the relevance of the models, interpret and generalize results, or even determine if the research followed best-practice scientific and ethical standards.

We performed a cross-sectional survey [6] of studies of solid-tumour oncology mouse models to evaluate the reporting of items specifically related to animal care and welfare, animal-related cancer aetiology, and endpoint expression. We confined searches to recent major cancer journals that explicitly endorsed the ARRIVE (Animal Research Reporting: In Vivo Experiments) reporting guidelines [7, 8] because these guidelines provide an objective benchmark for quality reporting expectations [9]. The primary intent of this survey was neither to synthesize evidence (as with systematic reviews), nor perform a complete evaluation of adherence to all reporting items identified by ARRIVE. Instead, the main objective of this investigation was to provide a prevalence snapshot of commonly-overlooked reproducibility 'risk factors' specifically associated with animal use, humane care, and welfare. To ensure rigorous review, we followed standards for conduct and reporting of scoping reviews (The PRISMA Extension for Scoping Reviews (PRISMA-ScR) [10]. Oncology studies utilizing mouse models were evaluated for reporting of items in five key animal-specific domains (ethical oversight assurance, animal signalment, husbandry, welfare, and euthanasia), and evaluated for the extent of reporting compliance and major gaps. We also evaluated reporting of simple study validity items (sample size, sample size justification, and bias minimisation) to enable comparison with other, more general, reviews. We discuss how reporting gaps identified in this survey limit the utility and translatability of mouse oncology models, and provide a list of targeted recommendations to improve the quality of these studies.

Materials & methods

Eligibility and screening

Data were extracted from 400 articles in 20 oncology research journals representing six publishing groups (**Fig 1**). The study was purposefully restricted to a single year (2020). A total of 284 journals related to oncology research were identified and screened. Journals were selected if the primary focus was on preclinical studies involving animal use, and if they explicitly endorsed ARRIVE reporting guidelines in the Instructions to Authors [7, 8]. Journals were excluded if subject matter was predominantly or exclusively clinical and/or molecular, as indicated by both electronic search on the terms ((mouse OR mice OR murine) OR preclinical OR animal) and visual search of article titles, abstracts, and main text in each journal. Impact factors ranged from 2.97 to 26.5, as determined from the 2020 Journal Citation Reports (Clarivate Web of Science). We selected the first 20 articles in each journal for the year 2020 that described original experimental research involving mouse oncology models. Clinical or
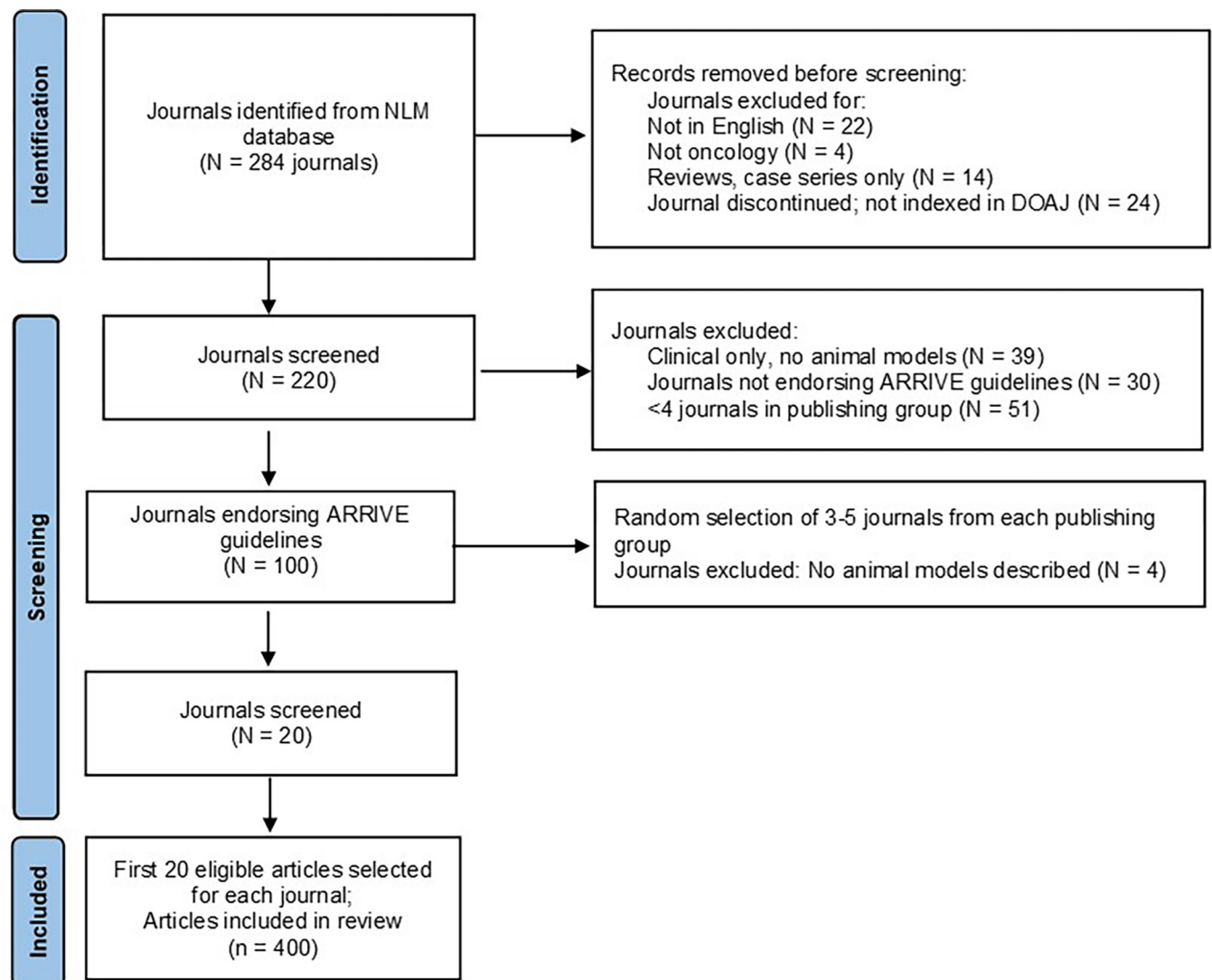

Fig 1. Flow diagram for journal identification, article selection, screening, inclusion, and exclusion.

https://doi.org/10.1371/journal.pone.0274738.g001

epidemiological studies, *in vitro* studies, letters to the editor, conference abstracts, and reviews were not included. Journal and article selection processes are described in more detail in the **S1 File: Supplementary Methods.**

Data

Both authors independently screened each article by examining Materials and Methods, Results, and article supplementary files (if provided) for reference to experiments involving mice. Key reporting items in five domains (ethical oversight assurance, animal signalment, animal husbandry, welfare-related items, and euthanasia), identified from the relevant sections of ARRIVE guidelines [7, 8] and itemised in a checklist (**S1 File: Supplementary Methods**) were scored as reported (1) or not reported (0). Only information included in the main text and supplementary methods was included. Information reported in figure legends but not mentioned elsewhere in the text was not considered as reported. The subset of articles that reported measurements of subcutaneous tumour volume as an experimental outcome (n = 290) were scored for tumour burden metrics reporting (method, sites, volume, maximum tumour size, time to maximum tumour size). All articles were scored for simple study validity metrics (total sample size, sample size justification, randomisation, blinding; **S1 File: Supplementary Methods**).

Both authors scored all articles individually in separate spreadsheets (Microsoft Excel 2019; Microsoft Corporation, Redmond, WA). Discrepancies between scored items were flagged electronically. Discrepant entries were then compared to information in the original article for correction if necessary, and remaining ambiguities or divergence resolved by consensus. Data were imported into SAS 9.4 (Windows 10PRO x64; SAS Institute Inc., Cary, NC) for analysis. Further details are given in **S1 File: Supplemental Methods.**

Statistical analyses

This was a descriptive study rather than a hypothesis-testing study. Data were itemized and summarized by counts and percentages. Patterns of reporting compliance for care and welfare items were analysed using two-level hierarchical generalised random-intercept models, with articles clustered within journal, no predictors, and dichotomous (binary) outcomes [11, 12]. There were no previously published estimates for expected between- and within-cluster variances. Therefore, sample sizes for journals and articles per journal were selected to give reasonable estimates and precision for model parameters [13]. Simulations have indicated that confidence intervals with approximately correct coverage rates and minimal downwards bias can be obtained with approximately 20 observations per cluster and at least 20 clusters [14]. Poor reporting of study validity items precluded formal analysis.

The binary response (yes/no) for each reporting item was modelled as $\eta_{ij} = \gamma_0 + u_{0j}$ where η_{ij} is the log odds of a given item being reported for article i in journal j, γ_0 is the random intercept component representing the log odds of an item being reported for a given journal, and u_{0j} is the journal-level error, with u_{0j} assumed to be normally distributed with mean 0 and variance τ_0. The probability of reporting compliance for each item was calculated as $\phi_{ij} = e^{\eta_{ij}}/(1 + e^{\eta_{ij}})$. The amount of variation in item reporting accounted for by journal membership was assessed by intraclass correlation coefficients (ICC). ICC is estimated as the proportion of variance that can be attributed to between-journal variation: $ICC = \sigma_{jour}^2/(\sigma_{jour}^2 + \sigma_e^2)$, where σ_{jour}^2 is level-2 or between-journal variation, estimated from the covariance parameter estimate, and σ_e^2 is the level-1 or article-level variance, estimated as $\pi^2/3$ for the standard logistic distribution. (The variance for a hierarchical generalized linear model with binary outcomes is directly determined by the population mean [11, 12]). The ICC can

also be interpreted as the correlation ρ of the response for any two articles selected at random from the same journal [15]. Models were fitted using Laplace estimation in SAS *proc glimmix* (SAS v.9.4, SAS Institute, Cary NC; **S1 File: Supplemental Methods;** [11]).

Results

General

The summary of reporting compliance for care and welfare items is given in **Table 1.** Count summaries by journal are given in the **S2 File: Supplemental Results**. Median reporting compliance was 23% (IQR 21, 27%) for all reporting items across journals in this survey. There was no apparent relationship between percent compliance and journal or journal impact factor (**Fig 2**). Unadjusted correlation for care and welfare items with journal impact factor was r = −0.05 (95% confidence interval −0.48, 0.40), and the correlation of validity items with impact factor was r = −0.18 (95% confidence interval −0.57, 0.29).

Reporting (**Table 1**) was highest for minimal ethical oversight statements (84%), nominal strain identification (91%) and animal age class (81%), and lowest for animal husbandry (0–10%), welfare (2–16%), and euthanasia (7%) items. The intraclass correlation coefficients (ICC) indicate how much of the total variation in the probability of reporting a specific item is accounted for by journal. ICCs were for the most part moderate to poor, reflecting rates of reporting that were either uniformly high (e.g. ethical oversight items) or uniformly poor (e.g. welfare items) across all journals. Items with ICC of zero reflect either almost no reporting at

Table 1. Summary of reporting probability ϕ (proportion of articles reporting a given item, corrected for journal membership) with 95% confidence intervals, and intraclass correlation (ICC), estimated from two-level hierarchical models of articles nested within journal.

|			Reporting Probability			
		Number of articles, *n*	ϕ	95% CI		ICC
Ethical oversight	Institutional approval	331	0.838	0.779	0.883	0.28
	Approval number	105	0.236	0.160	0.332	0.38
	Care and use guidelines	174	0.430	0.350	0.515	0.29
Signalment	Nominal strain identifiers	356	0.909	0.851	0.945	0.35
	Verifiable strain identifiers	34	0.045	0.018	0.106	0.64
	Source or derivation	290	0.745	0.659	0.816	0.33
	Age	325	0.813	0.766	0.853	0.24
	Sex	287	0.725	0.659	0.782	0.27
	Weight	45	0.082	0.043	0.149	0.45
Husbandry	Caging	16	0.028	0.011	0.071	0.41
	Number of mice per cage	14	0.018	0.005	0.058	0.59
	Enrichment	4	0.000	0.000	0.056	0.00
	Temperature	28	0.033	0.011	0.093	0.71
	Photoperiod	45	0.072	0.034	0.143	0.55
	Food and water access	47	0.096	0.058	0.156	0.36
	Feed type	16	0.021	0.006	0.066	0.60
	Acclimation	18	0.044	0.023	0.083	0.25
Welfare	Anaesthesia	59	0.127	0.082	0.192	0.33
	Analgesia	11	0.022	0.008	0.059	0.34
	Humane endpoints	65	0.163	0.129	0.202	0.00
Euthanasia	Method identified	78	0.135	0.066	0.255	0.72
	Not specified	213	0.530	0.419	0.637	0.37
	Not reported	108	0.261	0.202	0.329	0.27

https://doi.org/10.1371/journal.pone.0274738.t001
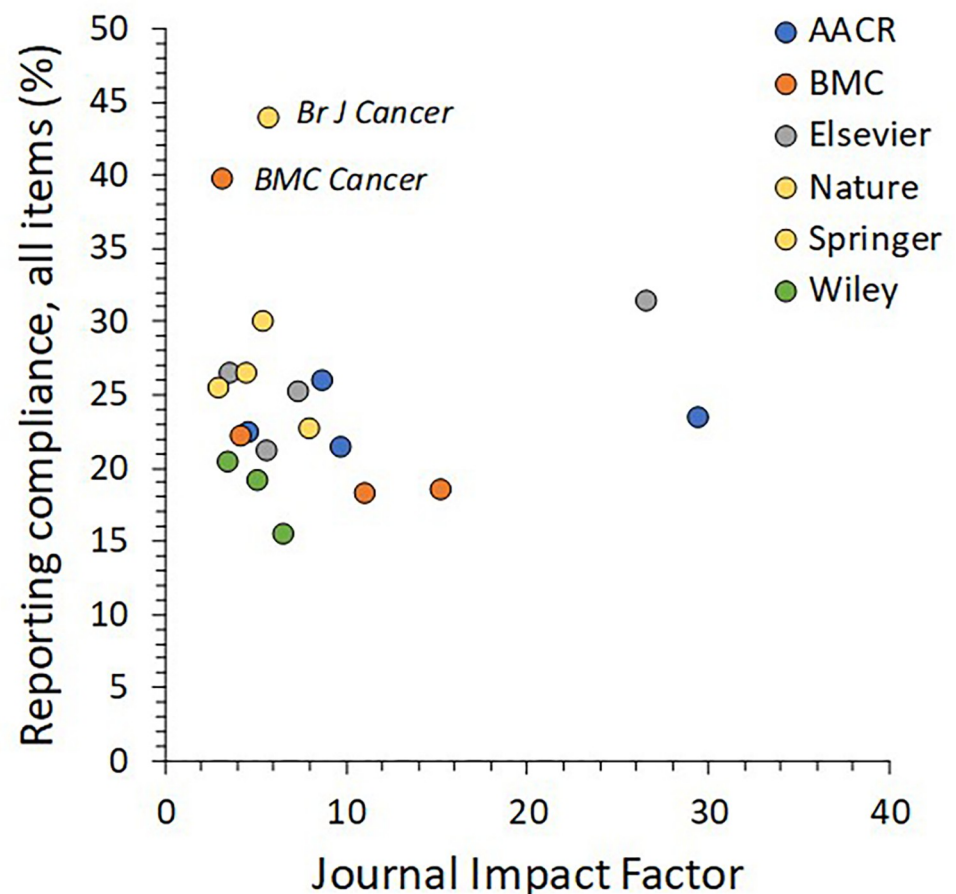
Fig 2. Association of journal impact factors and percent overall animal care & welfare reporting for mouse-based oncology studies in 20 journals.

https://doi.org/10.1371/journal.pone.0274738.g002

all (enrichment, n = 4/400 articles), or rare intermittent and inconsistent reporting across multiple journals (humane endpoints n = 65/400). Items with the highest ICC resulted from low overall frequency of reporting, resulting from high reporting concentrated in only one or two journals and poor reporting for all other journals. For example, 'temperature' has an overall reporting frequency of 3.3% (ϕ = 0.033) but an ICC of 0.71, reflecting 13 reports located in only two journals, and 15 reports in the remaining 18 journals (**S2 File**: **Supplemental Results**).

Ethical oversight

Approximately 84% (ϕ = 0.838) of articles explicitly reported that prior institutional approval was obtained, and 43% reported adherence to recognised and verifiable standards of humane animal care and use. However, fewer than one in four (24%) provided verifiable institutional protocol or license numbers, and 26/400 (7%) did not report any verifiable declaration of either institutional approval or care and use guidelines. One article claimed that their study "was not required to complete an ethical assessment", and one study stated that approval had been obtained "retrospectively". Approval numbers and oversight statements of six articles were identical to those reported in articles on different topics published in different journals

by unrelated research groups. Animal Welfare Assurance Numbers, which do not refer to individual animal use protocols but are granted to Public Health Service (PHS) awardee institutions, were reported by seven articles.

Animal signalment

Signalment is the complete description of the animal model itself, and should include unambiguous identification of strain, strain source (a recognised vendor, repository, or laboratory), age, sex, and if possible, body weight, as stipulated in the ARRIVE guidelines. For genetically modified strains, either reference to recognised official consensus identifiers (RRID, vendor stock or strain numbers), or a complete description of genotype derivation should be reported [16]. Although 91% of surveyed articles provided at least a nominal strain identification, only 75% provided a source or recognised vendor, and fewer than 5% of articles provided clear, complete and verifiable genotype identifiers or a sufficiently detailed description of breeding stock development. In 9% (44/400) of surveyed studies, strains used were either not identified, or only vague descriptors (such as "nude mice") were provided. The journal *Cancer Cell*, which uses a structured reporting format for methods (STAR-Methods, [17]) had the most thorough documentation of strain identifiers, with 12/20 articles supplying complete information.

The majority of research articles reported sex (73%) and age (81%). Nearly half (47%) reported using animals 4 to 6 weeks of age, although it was not clear if these were the ages at which animals were acquired, or if they were the ages at which experimental manipulations occurred. Body weights were reported for only 45/400 (11%) of surveyed studies.

Husbandry

Key husbandry information was poorly reported. Basic information on caging, cage density (number of mice per cage), and enrichment was almost never described (0–3%). Environmental variables temperature and photoperiod were also poorly reported (3% and 7% respectively).

Welfare

Descriptions of welfare-related assessments and procedures (analgesia and anaesthesia, postoperative and palliative care, and humane endpoints) were poorly reported. Anaesthesia and analgesia use were reported by 13% and 2% of articles respectively. Articles reporting use of specific agents rarely provided necessary information on methods of administration, dose, route, concentration, manufacturer, indications for use, and/or administration schedules. Less than 1% (3/400) papers described pre-emptive analgesia use, 2.5 (10/400) reported using post-operative analgesia, and 2% (8/400) reported use of opioids. Most studies reported monitoring experimental animals for days to weeks post-tumour induction and before euthanasia, and 52% reported some sort of "survival analysis" using methods for time to event data (e.g. Kaplan-Meier estimates, Cox proportional hazards regression). However, only 16% (65/400 studies) reported specific humane endpoints. No study reported direct assessments of pain-related behaviours or response to palliative care measures.

Tumour volume is commonly used to assess disease progression, tumorigenicity, and response to therapeutic intervention [18], and tumour burden is a critical humane endpoint [3]. Specific methodological details for tumours are not explicitly identified in the ARRIVE guidelines. Nevertheless, it should be apparent that all methods used to determine a key experimental endpoint should be reported in adequate detail, and specific guidelines for reporting tumour burden and humane endpoints have been available for over a decade [3]. Of the 290 articles describing subcutaneous tumour volume as endpoint, pertinent descriptors necessary

for volume determination were not reported by the majority of papers. Twenty-seven percent (78/290) did not report the anatomical site of tumour induction or gave only a vague description, 52% (152/290) did not report whether induction sites were unilateral or bilateral, 53% (152/290) did not report the measurement tool used (e.g. callipers), 29% (85/290) did not report the volume calculation formula, 72% (210/290) did not report the *a priori* maximum allowable tumour size for humane endpoint, and 59% (172/290) did not report the *a priori* maximum duration allowable for tumour growth prior to euthanasia. Inspection of results and figures suggested that at least 39% of reported studies (114/290) allowed tumours exceed the recommended limit of 1500 mm^3 without scientific justification [3], 9% (46/290) showed animals with tumours exceeding 3000 mm^3, and 8 articles reported tumour sizes exceeding 6000 mm^3. Another 16% (46/290) either did not report volumes at all (although describing the methods of doing so), or reported only 'relative' or 'normalised' volumes or nonstandard metrics which could not be assessed. There was also considerable variety in methods of calculating tumour volume. Nearly half (49%, 141/290) of surveyed studies measured tumour size with external callipers and calculated volume as a quadrangular prism in two dimensions (length x $width^2$/2). However, the remainder described use of up to 11 different formulae or variations. No paper reported determination of associated measurement errors or intra-observer variation.

Euthanasia

Euthanasia methods were explicitly identified in only 14% of surveyed articles. Only two journals (*BMC Cancer*, *British Journal of Cancer)* consistently reported euthanasia methods (90%; 36/40). Although over half (53%) all surveyed articles reported that animals were euthanized, no methods were identified or described in these papers. The remaining 27% of articles did not report any method of animal disposition before tissue harvest.

Study validity

Although all studies reported results of null hypothesis statistical tests, few studies reported verifiable information for total numbers of animals used, formal sample size justification, or bias minimisation methods (randomisation, blinding). Only 15% (59/400) articles provided 'total' sample sizes, and 31% (124/400) gave a sample size per intervention arm. However, these numbers are likely an under-estimate of the number of animals used, as animal loss due to attrition and discarded experiments was not recorded, nor was it clear how many whole-animal experiments were actually conducted, or even how many intervention arms were tested. Formal sample size justification using power calculations was claimed by 7 papers (2%), although descriptions were too incomplete to allow verification. Another 16 papers provided other unverifiable forms of sample size 'justification', based on 'previous experience', 'mouse availability and feasibility', numbers 'as small as possible to produce valid results' or the 'number used to obtain statistically significant results'. 'Randomisation' was claimed by 41% (165/400), with apparent stratification on tumour volume, animal weight, or age by 7 (<2%). Only 4 articles described using software or a random numbers table, and none described the randomisation method or the unit of randomisation. Two articles described as 'random' allocation that was sequential or alternating, respectively. Blinding was mentioned by 16 articles (4%), but none described how concealment was performed.

Discussion

Mice are the 'invisible actors' in much pre-clinical oncology research. In spite of explicit endorsement of the ARRIVE reporting guidelines by all journals in this survey, reporting of

animal-related information was inadequate. Descriptions of experimental techniques and procedures were emphasised at the expense of critical animal-related detail, and many details essential for assessing both study reliability and animal welfare were not reported. The widespread failure to report these details represent significant methodological omissions in preclinical oncology research.

Unfortunately, the results of this survey are consistent with those of other recent reviews that have found poor reporting compliance for other research specialties [19, 20]. There is increasing concern that the lack of reproducibility of much animal-based research is directly related to poor methodological documentation of critical information both 'inherent to the animals' (such as strain, age, sex etc.), but also those 'extrinsic factors of the animals' environment . . . that systematically influence the experimental outcomes' [21–23]. Without these details, much of the evidence claiming translation potential of mouse-based oncology models will be suspect. Complete and accurate reporting of experimental details is crucial for assessing model relevance, potential sources of variation and model disparity, translation potential, and (not least) if animal-based research has been conducted in compliance with best-practice animal welfare standards.

Ethical oversight of animal research is a fundamental research requirement. All journals in the current survey expressly stipulated that prospective studies required approval from the relevant institutional oversight committee before research animals were obtained or used. Therefore, it was both disappointing and unexpected that unambiguous and verifiable statements of institutional approval showed much less than 100% compliance, nearly one in ten studies failed to report any verifiable ethical oversight information at all, and some provided information that was demonstrably false. Poor animal care and use is poor science. Without verifiable ethical oversight information, it is impossible to tell if ethical review of the animal experiments was actually performed, if studies were conducted under appropriate ethical oversight, or if the work followed best practices for humane care and use. Plagiarised or false ethical oversight information is research misconduct.

Standardized, genetically defined mouse strains and stocks are primary biomedical research tools. However, in this survey, <10% of studies reported verifiable strain descriptions or used standardised nomenclature. This is of concern, because mice, and especially inbred strains, are subject to both obvious and quiet genetic mutations with each round of breeding [16, 24–26]. These mutations can be due to genetic drift with differential fixation, or genetic contamination resulting from breeding colony mismanagement [27]. Quiet mutations are the most problematic because they do not result in a readily visible phenotype, and can go undetected unless genetic stability testing is performed routinely [16]. Sub-strains of inbred lines produced by different commercial vendors, or continuous in-house breeding programs, may differ both genetically and phenotypically, thus contributing to variation and compromising data interpretation [28]. The scientific community must follow basic guidelines for breeding and describing research animals, such as those endorsed by FELASA (Federation of European Laboratory Animal Science Associations) [16]. Meticulous and accurate identification of mouse lines used is essential to ensure the assumed genetic model is in fact the correct model for purpose, facilitate scientific communication, and improve reproducibility. Without specific information on genetic background and strain derivation, it will not be possible to assess model relevance, identify appropriate controls, or interpret and generalize the results in a meaningful way [28, 29].

Descriptions of housing and husbandry practices are frequently overlooked in methods reporting, and in this survey were rarely reported. These omissions are of concern for study reproducibility in general, because housing and husbandry conditions can have profound effects on health and welfare of mice, and can be a cause of phenotypic variation. More

specifically for oncology studies, housing and environmental conditions can greatly affect rates of tumour induction, invasion or remission, and response to test interventions. For example, both tumour formation and development in mice are affected by solitary versus social housing. Social isolation stress enhances tumour invasion, metastasis [30–34], tumour growth [34–44], and gene expression, and attenuates response to chemotherapy [35]. Compared to solitary animals, group-housed animals typically have smaller tumours and increased rates of tumour regression or rejection [40, 45, 46]. Additional environmental factors affecting tumour kinetics and gene expression include housing on ventilated racks [47–55], temperature (heat or cold) stress [56–62], bedding type [63], and enrichment [62, 64, 65]. Handling methods [47, 66] and the diet fed [67–72] also contribute to animal stress, and therefore may be expected to influence tumour growth and metastasis.

Results of this survey support prior observations [19, 73, 74] that preclinical research studies do not consistently report use of anaesthesia, analgesia, or other pain control measures. Even if surgery is not performed, general anaesthesia is generally used as a restraint agent in imaging studies. Choice of anaesthetic agents can introduce considerable experimental artefact and must be identified and justified [3]. Further, it is an ethical imperative to minimize pain and distress of animals used in invasive research. Pain management can be a methodological challenge if anaesthesia or analgesia agents have the potential to affect experimental endpoints, such as engraftment, tumour growth, or metastasis [75–77]. Thus, both cancer pain and methods of pain control have the potential to act as meaningful confounding factors [74]. Nevertheless, there is no good reason why responsible pain alleviation cannot be used [73]. Pain is a common clinical effect of many cancer types, and pain management is an integral part of human and veterinary oncology practice. Most cancer studies are conducted under the assumption that major morbidity and mortality will result from tumour progression without intervention. Analgesic use promotes welfare in animal oncology models by sustaining tumour growth to predetermined experimental endpoints without undue animal suffering [78]. Pre-emptive, perioperative and follow-up administration of pain relief measures are required for major and/or multiple survival surgeries, and studies with long post-injury monitoring periods. Analgesia should be the default for research protocols, and there should be very high scientific and ethical bar for withholding analgesia.

Given that many oncology studies have the potential for animal suffering and death, it is of major concern that information on welfare monitoring, humane endpoints, and euthanasia was reported so infrequently. Humane endpoints must be defined for so-called "survival studies" because death as an endpoint is discouraged by most reputable ethical oversight bodies. Failure to euthanize animals at predefined humane endpoints can lead to significant animal suffering and poor welfare. Potential adverse events and clinical signs associated with the therapeutic compounds under test should also be categorised *a priori*, identified and reported, especially if long-term study goals include translation.

Specification of humane endpoints should prioritise clear descriptions of methods for determining tumour burden, and predefined limits to maximum permissible tumour burden and duration of tumour growth. Because tumours vary in size and aggressiveness depending on the cancer type and location, ethical expectations are that protocols must also include clearly defined study-specific humane endpoint criteria, descriptions of monitoring frequency, and methods used to minimise suffering of the animals. More conservative tumour burden limits must also be considered if multiple tumours are present. Rigorous specification of these key metrics is necessary, both as reliable measures of tumorigenesis and intervention effects, and as humane endpoint indicators. Published consensus guidelines have been available for over a decade that specify limits to tumour size consistent with humane use and study validity (e.g. [3]), and animal ethics oversight committees usually have clear specifications for

maximum permissible tumour size. It was therefore disappointing that this survey showed that these items were poorly reported.

Evaluation of tumorigenesis data is further confounded by lack of consistent standards for tumour volume determinations. External calliper measurements (the most commonly used method) are prone to major systematic biases and observer variability [18, 79]. These problems are exacerbated by different methods of estimating tumour volume from linear measurements, which may result in considerable under- or over -estimation of tumour sizes. Because certain cancers have the potential for explosive tumour growth, measurement frequency should be tailored to the specific cancer type to avoid tumours exceeding allowable limits between measurement intervals. The inappropriate reporting of 'relative' or 'normalised' tumour volumes, alternative metrics or methods of calculating volume that do not relate reliably to tumour burden, plus failure to define maximum time and burden limits, also contribute to lack of transparency and oversight.

Methods of euthanasia can induce large differences in protein, metabolite, and biomarker expression, depending on both agent and pre-euthanasia versus post-euthanasia timing of tissue collection [80–82]. Our finding that 80% (321/400) of articles in this survey did not report any verifiable euthanasia methods at all is also concerning, as it is therefore impossible to compare results based on tissue harvest data.

Previous reviews have reported uniformly disappointing results for reporting of study validity and risk of bias items [10, 19, 21, 83]. This survey showed shared the (sadly) common features of inadequate sample size reporting, inappropriate justification, and poor understanding of basic concepts involved with bias minimisation. The extent of claimed randomisation observed in this study is probably greatly exaggerated, as it is likely that many investigators conflate 'random' with 'haphazard' or 'unplanned', and no study provide sufficient detail to indicate if the appropriate units of analysis were used in subsequent hypothesis tests. There was no indication that journal impact factor made much, if any, difference to risk of bias in published articles.

Limitations of this study include the potential for selection bias and lack of quality appraisal of the included studies [7, 84]. We selected only oncology journals from the larger established publishing groups, excluded those published in a language other than English, and included only those that explicitly endorsed the ARRIVE guidelines. Because reporting guidelines represent the minimum information necessary for assessing research reliability [9], we reasoned that journals endorsing such guidelines would provide a compliance benchmark for other journals not following such guidelines. However, this was not the case. Reporting compliance overall was extremely poor, with major methodological reporting gaps and no apparent relationship between rates of reporting and journal impact factor. Second, because this was a cross-sectional survey and not a formal systematic review, we did not evaluate the quality of evidence for individual studies, differences between cancer models, or validity of results [83]. Instead our goal was to determine patterns of reporting and reporting gaps for animal care and welfare items known to contribute to variability in response to both cancer induction methods and experimental interventions. We cannot rule out bias in the studies themselves. Perceived bias against publication of "negative" results may mean that investigators probably do not report all experiments with all animals, but only those with "significant" findings [83].

A recent survey of preclinical investigators indicated that the major reason for failure to report ARRIVE items in research reports was because those items were not considered 'important' or 'necessary' [20]. The widespread omissions noted in this survey indicate that animal-related information, although specifically singled out in the ARRIVE guidelines, definitely takes a back seat to that for other resources and procedures. Unfortunately, reporting quality has remained consistently poor across diverse research specialties and journals since the

guidelines were introduced [10, 19]. One factor contributing to the lack of improvement is the current publication fashion for the reporting of numerous diverse experiments in a single paper, presumably to indicate if results are 'robust'. This has resulted in highly information-dense reporting of the results for numerous individual experiments and incomplete methodology documentation, making studies difficult or impossible to review for substance [85]. A second factor is the widespread failure of journal and peer reviewers to actively enforce agreed-upon best-practice reporting standards [19, 20].

Investigators, journal editors, and peer reviewers need to put the 'mouse' back into mouse model-based research. Editorial and journal staff must be more actively involved in enforcing reporting standards [9, 20] and ensure that all relevant animal-based information (including details of ethical oversight) are described. Mandatory completion of ARRIVE or Structured, Transparent, Accessible Reporting (STAR Methods) checklists by the submitting authors has been documented to significantly improve scientific reporting when enforced by journals [86–88]. The reform of journal content in the direction of fewer, but better and more thoroughly documented and reported experiments should be prioritised. This would have the added advantage of reducing information overload and therefore the burden on reviewers, and possibly contribute to more thorough reviews [85].

Accountability in science is key to improving practices. It is a scientific imperative to ensure that models are relevant and translatable. It is an ethical imperative to minimize pain and distress in animals used in invasive research and that appropriate oversight guardrails are in place. If methodological substance is de-emphasised in favour of the narrative of results, preclinical oncology research will continue to be compromised.

Supporting information

S1 Checklist. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.
(PDF)

S1 File. Supplemental methods.
(DOCX)

S2 File. Supplemental results, S1-S6 Tables.
(DOCX)

S1 Data.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Elizabeth A. Nunamaker, Penny S. Reynolds.

Data curation: Elizabeth A. Nunamaker, Penny S. Reynolds.

Formal analysis: Penny S. Reynolds.

Investigation: Elizabeth A. Nunamaker, Penny S. Reynolds.

Methodology: Elizabeth A. Nunamaker, Penny S. Reynolds.

Validation: Elizabeth A. Nunamaker, Penny S. Reynolds.

Visualization: Elizabeth A. Nunamaker, Penny S. Reynolds.

Writing – original draft: Elizabeth A. Nunamaker, Penny S. Reynolds.

Writing – review & editing: Elizabeth A. Nunamaker, Penny S. Reynolds.

References

1. Cheon DJ, Orsulic S. Mouse models of cancer. Annu Rev Pathol. 2011; 6:95–119. https://doi.org/10.1146/annurev.pathol.3.121806.154244 PMID: 20936938

2. Cook N, Jodrell DI, Tuveson DA. Predictive in vivo animal models and translation to clinical trials. Drug Discov Today. 2012; 17(5–6):253–60. https://doi.org/10.1016/j.drudis.2012.02.003 PMID: 22493784

3. Workman P, Aboagye EO, Balkwill F, Balmain A, Bruder G, Chaplin DJ, et al. Guidelines for the welfare and use of animals in cancer research. Br J Cancer. 2010; 102(11):1555–77. https://doi.org/10.1038/sj.bjc.6605642 PMID: 20502460

4. Smith AJ, Lilley E. The role of the Three Rs in improving the planning and reproducibility of animal experiments. Animals (Basel). 2019; 9(11). https://doi.org/10.3390/ani9110975 PMID: 31739641

5. Garner JP, Gaskill BN, Weber EM, Ahloy-Dallaire J, Pritchett-Corning KR. Introducing Therioepistemology: the study of how knowledge is gained from animal research. Lab Anim (NY). 2017; 46(4):103–13. https://doi.org/10.1038/laban.1224 PMID: 28328885

6. Koffel JB, Rethlefsen ML. Reproducibility of search strategies is poor in systematic reviews published in high-impact pediatrics, cardiology and surgery journals: A cross-sectional study. PLoS One. 2016; 11(9):e0163309. https://doi.org/10.1371/journal.pone.0163309 PMID: 27669416

7. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: The ARRIVE guidelines for reporting animal research. PLoS Biol. 2010; 8(6):e1000412. https://doi.org/10.1371/journal.pbio.1000412 PMID: 20613859

8. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. PLoS Biol. 2020; 18(7):e3000410. https://doi.org/10.1371/journal.pbio.3000410 PMID: 32663219

9. Avey MT, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, et al. The devil is in the details: Incomplete reporting in preclinical animal research. PLoS One. 2016; 11(11):e0166733. https://doi.org/10.1371/journal.pone.0166733 PMID: 27855228

10. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol. 2014; 67(12):1291–4. https://doi.org/10.1016/j.jclinepi.2014.03.013 PMID: 25034198

11. Ene M, Leighton E, Blue G, Bell B. Multilevel models for categorical data using SAS PROC GLIMMIX: The basics. [1/5/21]. Available from: *https://support.sas.com/resources/papers/proceedings15/3430-2015.pdf*

12. Hox JJ, Moerbeek M, van de Schoot R. Chapter 6: The multilevel generalized linear model for dichotomous data and proportions. Multilevel Analysis: Techniques and Applications. Third ed: Taylor & Francis Group; 2018. p. 103–29.

13. Hox JJ, Moerbeek M, van de Schoot R. Chapter 12: Sample sizes and power analysis in multilevel regression. Multilevel Analysis: Techniques and Applications. Third ed: Taylor & Francis Group; 2018. p. 212–34.

14. Austin P. Estimating multilevel logistic regression models when the number of clusters is low: a comparison of different statistical software procedures. International Journal of Biostatistics. 2010; 6(1). https://doi.org/10.2202/1557-4679.1195 PMID: 20949128

15. Austin PC, Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. Stat Med. 2017; 36(20):3257–77. https://doi.org/10.1002/sim.7336 PMID: 28543517

16. Benavides F, Rülicke T, Prins JB, Bussell J, Scavizzi F, Cinelli P, et al. Genetic quality assurance and genetic monitoring of laboratory mice and rats: FELASA Working Group Report. Lab Anim. 2020; 54(2):135–48. https://doi.org/10.1177/0023677219867719 PMID: 31431136

17. Marcus E. A STAR Is Born. Cell. 2016; 166(5):1059–60. https://doi.org/10.1016/j.cell.2016.08.021 PMID: 27565332

18. Delgado-SanMartin J, Ehrhardt B, Paczkowski M, Hackett S, Smith A, Waraich W, et al. An innovative non-invasive technique for subcutaneous tumour measurements. PLoS One. 2019; 14(10):e0216690. https://doi.org/10.1371/journal.pone.0216690 PMID: 31609977

19. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVEd: Support for the ARRIVE (Animal Research: Reporting of in vivo Experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. PLoS One. 2018; 13(5):e0197882. https://doi.org/10.1371/journal.pone.0197882 PMID: 29795636

20. Hair K, Macleod MR, Sena ES, Collaboration I. A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). Res Integr Peer Rev. 2019; 4:12. https://doi.org/10.1186/s41073-019-0069-3 PMID: 31205756

21. Wold B, Tabak L. ACD Working Group on Enhancing Rigor, Transparency, and Translatability in Animal Research 2021. Available from: https://acd.od.nih.gov/working-groups/eprar.html.

22. Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Challenges for assessing replicability in preclinical cancer biology. Elife. 2021;10. https://doi.org/10.7554/eLife.67995 PMID: 34874008

23. Errington TM, Mathur M, Soderberg CK, Denis A, Perfito N, Iorns E, et al. Investigating the replicability of preclinical cancer biology. Elife. 2021;10. https://doi.org/10.7554/eLife.71601 PMID: 34874005

24. Stevens JC, Banks GT, Festing MF, Fisher EM. Quiet mutations in inbred strains of mice. Trends Mol Med. 2007; 13(12):512–9. https://doi.org/10.1016/j.molmed.2007.10.001 PMID: 17981508

25. Fahey JR, Katoh H, Malcolm R, Perez AV. The case for genetic monitoring of mice and rats used in biomedical research. Mamm Genome. 2013; 24(3–4):89–94. https://doi.org/10.1007/s00335-012-9444-9 PMID: 23314661

26. Casellas J, Varona L. Short communication: Effect of mutation age on genomic predictions. J Dairy Sci. 2011; 94(8):4224–9. https://doi.org/10.3168/jds.2011-4186 PMID: 21787959

27. Taft RA, Davisson M, Wiles MV. Know thy mouse. Trends Genet. 2006; 22(12):649–53. Epub 2006/09/26. https://doi.org/10.1016/j.tig.2006.09.010 PMID: 17007958

28. Mahajan VS, Demissie E, Mattoo H, Viswanadham V, Varki A, Morris R, et al. Striking immune phenotypes in gene-targeted mice are driven by a copy-number variant originating from a commercially available C57BL/6 strain. Cell Rep. 2016; 15(9):1901–9. https://doi.org/10.1016/j.celrep.2016.04.080 PMID: 27210752

29. Rülicke T, Montagutelli X, Pintado B, Thon R, Hedrich HJ, Group FW. FELASA guidelines for the production and nomenclature of transgenic rodents. Lab Anim. 2007; 41(3):301–11. https://doi.org/10.1258/002367707781282758 PMID: 17640457

30. Wu W, Yamaura T, Murakami K, Ogasawara M, Hayashi K, Murata J, et al. Involvement of TNF-alpha in enhancement of invasion and metastasis of colon 26-L5 carcinoma cells in mice by social isolation stress. Oncol Res. 1999; 11(10):461–9. PMID: 10850887

31. Wu W, Yamaura T, Murakami K, Murata J, Matsumoto K, Watanabe H, et al. Social isolation stress enhanced liver metastasis of murine colon 26-L5 carcinoma cells by suppressing immune responses in mice. Life Sci. 2000; 66(19):1827–38. https://doi.org/10.1016/s0024-3205(00)00506-3 PMID: 10809180

32. Bartolomucci A. Social stress, immune functions and disease in rodents. Front Neuroendocrinol. 2007; 28(1):28–49. Epub 2007/02/16. https://doi.org/10.1016/j.yfrne.2007.02.001 PMID: 17379284

33. Hoffman-Goetz L, Simpson JR, Arumugam Y. Impact of changes in housing condition on mouse natural killer cell activity. Physiol Behav. 1991; 49(3):657–60. https://doi.org/10.1016/0031-9384(91)90297-2 PMID: 2062948

34. Hermes GL, Delgado B, Tretiakova M, Cavigelli SA, Krausz T, Conzen SD, et al. Social isolation dysregulates endocrine and behavioral stress while increasing malignant burden of spontaneous mammary tumors. Proc Natl Acad Sci U S A. 2009; 106(52):22393–8. https://doi.org/10.1073/pnas.0910753106 PMID: 20018726

35. Weinberg J, Emerman JT. Effects of psychosocial stressors on mouse mammary tumor growth. Brain Behav Immun. 1989; 3(3):234–46. https://doi.org/10.1016/0889-1591(89)90039-1 PMID: 2611411

36. Wu W, Murata J, Murakami K, Yamaura T, Hayashi K, Saiki I. Social isolation stress augments angiogenesis induced by colon 26-L5 carcinoma cells in mice. Clin Exp Metastasis. 2000; 18(1):1–10. https://doi.org/10.1023/a:1026548715669 PMID: 11206831

37. Hasegawa H, Saiki I. Psychosocial stress augments tumor development through beta-adrenergic activation in mice. Jpn J Cancer Res. 2002; 93(7):729–35. https://doi.org/10.1111/j.1349-7006.2002.tb01313.x PMID: 12149137

38. Madden KS, Szpunar MJ, Brown EB. Early impact of social isolation and breast tumor progression in mice. Brain Behav Immun. 2013; 30 Suppl:S135–41. https://doi.org/10.1016/j.bbi.2012.05.003 PMID: 22610067

39. Bartolomucci A, Palanza P, Sacerdote P, Ceresini G, Chirieleison A, Panerai AE, et al. Individual housing induces altered immuno-endocrine responses to psychological stress in male mice. Psychoneuroendocrinology. 2003; 28(4):540–58. https://doi.org/10.1016/s0306-4530(02)00039-2 PMID: 12689611

40. Grimm MS, Emerman JT, Weinberg J. Effects of social housing condition and behavior on growth of the Shionogi mouse mammary carcinoma. Physiol Behav. 1996; 59(4–5):633–42. https://doi.org/10.1016/0031-9384(95)02126-4 PMID: 8778846

41. Williams JB, Pang D, Delgado B, Kocherginsky M, Tretiakova M, Krausz T, et al. A model of gene-environment interaction reveals altered mammary gland gene expression and increased tumor growth following social isolation. Cancer Prev Res (Phila). 2009; 2(10):850–61. https://doi.org/10.1158/1940-6207.CAPR-08-0238 PMID: 19789294

42. Strange KS, Kerr LR, Andrews HN, Emerman JT, Weinberg J. Psychosocial stressors and mammary tumor growth: an animal model. Neurotoxicol Teratol. 2000; 22(1):89–102. https://doi.org/10.1016/s0892-0362(99)00049-5 PMID: 10642118

43. Hilakivi-Clarke L, Dickson RB. Stress influence on development of hepatocellular tumors in transgenic mice overexpressing TGF alpha. Acta Oncol. 1995; 34(7):907–12. https://doi.org/10.3109/02841869509127203 PMID: 7492379

44. Kerr LR, Wilkinson DA, Emerman JT, Weinberg J. Interactive effects of psychosocial stressors and gender on mouse mammary tumor growth. Physiol Behav. 1999; 66(2):277–84. https://doi.org/10.1016/s0031-9384(98)00296-0 PMID: 10336154

45. Kerr LR, Grimm MS, Silva WA, Weinberg J, Emerman JT. Effects of social housing condition on the response of the Shionogi mouse mammary carcinoma (SC115) to chemotherapy. Cancer Res. 1997; 57(6):1124–8. PMID: 9067282

46. Kerr LR, Hundal R, Silva WA, Emerman JT, Weinberg J. Effects of social housing condition on chemotherapeutic efficacy in a Shionogi carcinoma (SC115) mouse tumor model: influences of temporal factors, tumor size, and tumor growth rate. Psychosom Med. 2001; 63(6):973–84. https://doi.org/10.1097/00006842-200111000-00017 PMID: 11719637

47. Giraldi T, Perissin L, Zorzet S, Piccini P, Rapozzi V. Effects of stress on tumor growth and metastasis in mice bearing Lewis lung carcinoma. Eur J Cancer Clin Oncol. 1989; 25(11):1583–8. https://doi.org/10.1016/0277-5379(89)90301-5 PMID: 2591450

48. David JM, Knowles S, Lamkin DM, Stout DB. Individually ventilated cages impose cold stress on laboratory mice: a source of systemic experimental variability. J Am Assoc Lab Anim Sci. 2013; 52(6):738–44. PMID: 24351762

49. Åhlgren J, Voikar V. Housing mice in the individually ventilated or open cages-Does it matter for behavioral phenotype? Genes Brain Behav. 2019; 18(7):e12564. Epub 2019/03/28. https://doi.org/10.1111/gbb.12564 PMID: 30848040

50. Burman O, Buccarello L, Redaelli V, Cervo L. The effect of two different individually ventilated cage systems on anxiety-related behaviour and welfare in two strains of laboratory mouse. Physiol Behav. 2014; 124:92–9. https://doi.org/10.1016/j.physbeh.2013.10.019 PMID: 24184492

51. Kallnik M, Elvert R, Ehrhardt N, Kissling D, Mahabir E, Welzl G, et al. Impact of IVC housing on emotionality and fear learning in male C3HeB/FeJ and C57BL/6J mice. Mamm Genome. 2007; 18(3):173–86. https://doi.org/10.1007/s00335-007-9002-z PMID: 17431719

52. Logge W, Kingham J, Karl T. Behavioural consequences of IVC cages on male and female C57BL/6J mice. Neuroscience. 2013; 237:285–93. https://doi.org/10.1016/j.neuroscience.2013.02.012 PMID: 23415791

53. Mineur YS, Crusio WE. Behavioral effects of ventilated micro-environment housing in three inbred mouse strains. Physiol Behav. 2009; 97(3–4):334–40. https://doi.org/10.1016/j.physbeh.2009.02.039 PMID: 19281831

54. Pasquarelli N, Voehringer P, Henke J, Ferger B. Effect of a change in housing conditions on body weight, behavior and brain neurotransmitters in male C57BL/6J mice. Behav Brain Res. 2017; 333:35–42. https://doi.org/10.1016/j.bbr.2017.06.018 PMID: 28625548

55. Polissidis A, Zelelak S, Nikita M, Alexakos P, Stasinopoulou M, Kakazanis ZI, et al. Assessing the exploratory and anxiety-related behaviors of mice. Do different caging systems affect the outcome of behavioral tests? Physiol Behav. 2017; 177:68–73. https://doi.org/10.1016/j.physbeh.2017.04.009 PMID: 28412281

56. Eng JW, Reed CB, Kokolus KM, Pitoniak R, Utley A, Bucsek MJ, et al. Housing temperature-induced stress drives therapeutic resistance in murine tumour models through β2-adrenergic receptor activation. Nat Commun. 2015; 6:6426. https://doi.org/10.1038/ncomms7426 PMID: 25756236

57. Hylander BL, Repasky EA. Thermoneutrality, mice, and cancer: A heated opinion. Trends Cancer. 2016; 2(4):166–75. https://doi.org/10.1016/j.trecan.2016.03.005 PMID: 28741570

58. Messmer MN, Kokolus KM, Eng JW, Abrams SI, Repasky EA. Mild cold-stress depresses immune responses: Implications for cancer models involving laboratory mice. Bioessays. 2014; 36(9):884–91. https://doi.org/10.1002/bies.201400066 PMID: 25066924

59. Yamamoto H, Fujii K, Hayakawa T. Inhibitory effect of cold stress against acetaminophen-induced hepatic injury in B6C3F1 and ICR mice. Toxicol Lett. 1995; 81(2–3):125–30. https://doi.org/10.1016/0378-4274(95)03424-2

60. Yamamoto H, Fujii K, Hayakawa T. Inhibitory effect of cold stress on lung tumours induced by 7,12-dimethylbenz[a]anthracene in mice. J Cancer Res Clin Oncol. 1995; 121(7):393–6. https://doi.org/10.1007/BF01212944 PMID: 7635867

61. Gordon CJ, Aydin C, Repasky EA, Kokolus KM, Dheyongera G, Johnstone AF. Behaviorally mediated, warm adaptation: a physiological strategy when mice behaviorally thermoregulate. J Therm Biol. 2014; 44:41–6. https://doi.org/10.1016/j.jtherbio.2014.06.006 PMID: 25086972

62. Johnson JS, Taylor DJ, Green AR, Gaskill BN. Effects of nesting material on energy homeostasis in BALB/cAnNCrl, C57BL/6NCrl, and Crl:CD1(ICR) mice housed at 20°C. J Am Assoc Lab Anim Sci. 2017; 56(3):254–9. PMID: 28535860

63. Markaverich B, Mani S, Alejandro MA, Mitchell A, Markaverich D, Brown T, et al. A novel endocrine-disrupting agent in corn with mitogenic activity in human breast and prostatic cancer cells. Environ Health Perspect. 2002; 110(2):169–77. https://doi.org/10.1289/ehp.02110169 PMID: 11836146

64. Li G, Gan Y, Fan Y, Wu Y, Lin H, Song Y, et al. Enriched environment inhibits mouse pancreatic cancer growth and down-regulates the expression of mitochondria-related genes in cancer cells. Sci Rep. 2015; 5:7856. https://doi.org/10.1038/srep07856 PMID: 25598223

65. Rabadán R, Ramos-Campos M, Redolat R, Mesa-Gresa P. Physical activity and environmental enrichment: Behavioural effects of exposure to different housing conditions in mice. Acta Neurobiol Exp (Wars). 2019; 79(4):374–85. PMID: 31885394

66. Mertens S, Vogt MA, Gass P, Palme R, Hiebl B, Chourbaji S. Effect of three different forms of handling on the variation of aggression-associated parameters in individually and group-housed male C57BL/6NCrl mice. PLoS One. 2019; 14(4):e0215367. https://doi.org/10.1371/journal.pone.0215367 PMID: 30978250

67. Lv M, Zhu X, Wang H, Wang F, Guan W. Roles of caloric restriction, ketogenic diet and intermittent fasting during initiation, progression and metastasis of cancer in animal models: a systematic review and meta-analysis. PLoS One. 2014; 9(12):e115147. https://doi.org/10.1371/journal.pone.0115147 PMID: 25502434

68. Klement RJ, Champ CE, Otto C, Kämmerer U. Anti-tumor effects of ketogenic diets in mice: A meta-analysis. PLoS One. 2016; 11(5):e0155050. https://doi.org/10.1371/journal.pone.0155050 PMID: 27159218

69. Klement RJ, Fink MK. Dietary and pharmacological modification of the insulin/IGF-1 system: exploiting the full repertoire against cancer. Oncogenesis. 2016; 5:e193. https://doi.org/10.1038/oncsis.2016.2 PMID: 26878387

70. Bonorden MJ, Rogozina OP, Kluczny CM, Grossmann ME, Grande JP, Lokshin A, et al. Cross-sectional analysis of intermittent versus chronic caloric restriction in the TRAMP mouse. Prostate. 2009; 69 (3):317–26. https://doi.org/10.1002/pros.20878 PMID: 19016490

71. Bonorden MJ, Rogozina OP, Kluczny CM, Grossmann ME, Grambsch PL, Grande JP, et al. Intermittent calorie restriction delays prostate tumor detection and increases survival time in TRAMP mice. Nutr Cancer. 2009; 61(2):265–75. https://doi.org/10.1080/01635580802419798 PMID: 19235043

72. Rogozina OP, Bonorden MJ, Grande JP, Cleary MP. Serum insulin-like growth factor-I and mammary tumor development in ad libitum-fed, chronic calorie-restricted, and intermittent calorie-restricted MMTV-TGF-alpha mice. Cancer Prev Res (Phila). 2009; 2(8):712–9. https://doi.org/10.1158/1940-6207.CAPR-09-0028 PMID: 19654106

73. Carbone L, Austin J. Pain and laboratory animals: Publication practices for better data reproducibility and better animal welfare. PLoS One. 2016; 11(5):e0155001. https://doi.org/10.1371/journal.pone.0155001 PMID: 27171143

74. Taylor DK. Influence of pain and analgesia on cancer research studies. Comp Med. 2019; 69(6):501–9. https://doi.org/10.30802/AALAS-CM-19-000002 PMID: 31315692

75. Franchi S, Panerai AE, Sacerdote P. Buprenorphine ameliorates the effect of surgery on hypothalamus-pituitary-adrenal axis, natural killer cell activity and metastatic colonization in rats in comparison with morphine or fentanyl treatment. Brain Behav Immun. 2007; 21(6):767–74. https://doi.org/10.1016/j.bbi.2007.01.001 PMID: 17291715

76. Bratcher NA, Frost DJ, Hickson J, Huang X, Medina LM, Oleksijew A, et al. Effects of buprenorphine in a preclinical orthotopic tumor model of ovarian carcinoma in female CB17 SCID mice. J Am Assoc Lab Anim Sci. 2019; 58(5):583–8. https://doi.org/10.30802/AALAS-JAALAS-18-000127 PMID: 31412976

77. Husmann K, Arlt MJ, Jirkof P, Arras M, Born W, Fuchs B. Primary tumour growth in an orthotopic osteo-sarcoma mouse model is not influenced by analgesic treatment with buprenorphine and meloxicam. Lab Anim. 2015; 49(4):284–93. https://doi.org/10.1177/0023677215570989 PMID: 25650386

78. Lofgren J, Miller AL, Lee CCS, Bradshaw C, Flecknell P, Roughan J. Analgesics promote welfare and sustain tumour growth in orthotopic 4T1 and B16 mouse cancer models. Lab Anim. 2018; 52(4):351–64. https://doi.org/10.1177/0023677217739934 PMID: 29207902

79. Jensen MM, Jørgensen JT, Binderup T, Kjaer A. Tumor volume in subcutaneous mouse xenografts measured by microCT is more accurate and reproducible than determined by 18F-FDG-microPET or external caliper. BMC Med Imaging. 2008; 8:16. https://doi.org/10.1186/1471-2342-8-16 PMID: 18925932

80. Overmyer KA, Thonusin C, Qi NR, Burant CF, Evans CR. Impact of anesthesia and euthanasia on metabolomics of mammalian tissues: studies in a C57BL/6J mouse model. PLoS One. 2015; 10(2): e0117232. https://doi.org/10.1371/journal.pone.0117232 PMID: 25658945

81. Traslavina RP, King EJ, Loar AS, Riedel ER, Garvey MS, Ricart-Arbona R, et al. Euthanasia by CO_2 inhalation affects potassium levels in mice. J Am Assoc Lab Anim Sci. 2010; 49(3):316–22. PMID: 20587163

82. Pecaut MJ, Smith AL, Jones TA, Gridley DS. Modification of immunologic and hematologic variables by method of CO2 euthanasia. Comp Med. 2000; 50(6):595–602. PMID: 11200564

83. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A focus for improvement. PLoS Biol. 2015; 13(10):e1002273. https://doi.org/10.1371/journal.pbio.1002273 PMID: 26460723

84. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods. 2014; 5(4):371–85. https://doi.org/10.1002/jrsm.1123 PMID: 26052958

85. Amaral OB, Neves K. Reproducibility: Expect less of the scientific paper. Nature. 2021; 597(7876):329–31. https://doi.org/10.1038/d41586-021-02486-7 PMID: 34526702

86. Plint A, Moher D, Morrison A, Schulz K, Altman D, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Medical Journal of Australia. 2006; 185(5):263–7. https://doi.org/10.5694/j.1326-5377.2006.tb00557.x PMID: 16948622

87. Byrnes MC, Schuerer DJ, Schallom ME, Sona CS, Mazuski JE, Taylor BE, et al. Implementation of a mandatory checklist of protocols and objectives improves compliance with a wide range of evidence-based intensive care unit practices. Crit Care Med. 2009; 37(10):2775–81. https://doi.org/10.1097/CCM.0b013e3181a96379 PMID: 19581803

88. Han S, Olonisakin TF, Pribis JP, Zupetic J, Yoon JH, Holleran KM, et al. A checklist is associated with increased quality of reporting preclinical biomedical research: A systematic review. PLoS One. 2017; 12 (9):e0183591. https://doi.org/10.1371/journal.pone.0183591 PMID: 28902887