RESEARCH ARTICLE

Animal models in preclinical metastatic breast cancer immunotherapy research: A systematic review and meta-analysis of efficacy outcomes

Yalda Mirzaei[1], **Martina Hüffel**[1], **Sarah McCann**[2], **Alexandra Bannach-Brown**[2], **René H. Tolba**[1], **Julia Steitz**[1]*

1 Institute for Laboratory Animal Science, Uniklinik RWTH Aachen, Aachen, Germany, **2** Berlin Institute of Health at Charité (BIH), BIH QUEST Center for Responsible Research, Berlin, Germany

* jsteitz@ukaachen.de

Abstract

Breast cancer, particularly metastatic breast cancer (MBC), presents aggressive clinical challenges with limited treatment success. Immunotherapy has emerged as a promising approach, however, discrepancies between preclinical animal models and human cancers complicate translation to clinical outcomes. This systematic review and meta-analysis evaluated the effect of immunotherapy on primary and metastatic tumor regression in animal models of MBC and assessed the models' appropriateness and reproducibility to improve future preclinical study design. Following a preregistered protocol in PROSPERO (CRD42021207033), we conducted searches in MEDLINE, Embase, and Web of Science databases, yielding 2255 studies for title/abstract screening and 108 studies included after full-text screening. All included studies used mouse models, assessing primary outcomes through tumor volume or weight and metastatic outcomes via nodule count or bioluminescence. Only 14% of studies fully reported experimental animal characteristics, and 43% provided detailed experimental procedures. Of 105 articles (293 comparisons) included in the meta-analysis, pooled effect sizes indicated significant reductions in both primary and metastatic tumors. However, high heterogeneity across studies and wide prediction intervals suggested substantial variability in model responses to immunotherapy. Univariable and multivariable meta-regressions failed to significantly explain this heterogeneity, suggesting additional factors may influence outcomes. Trim-and-fill and Egger's regression tests indicated funnel plot asymmetry, implying potential publication bias and small study effects. While our analysis demonstrated positive effects of immunotherapy on MBC and highlighted variability in animal tumor models, addressing model-related heterogeneity and enhancing methodological transparency are essential to improve reproducibility and clinical translatability.

OPEN ACCESS

Citation: Mirzaei Y, Hüffel M, McCann S, Bannach-Brown A, Tolba RH, Steitz J (2025) Animal models in preclinical metastatic breast cancer immunotherapy research: A systematic review and meta-analysis of efficacy outcomes. PLoS One 20(5): e0322876. https://doi.org/10.1371/journal.pone.0322876

Editor: Mohamed Abdelkarim, Faculty of Medicine of Tunis, TUNISIA

Received: November 26, 2024

Accepted: March 28, 2025

Published: May 7, 2025

Copyright: © 2025 Mirzaei et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The data and analytic code for this study are publicly available on the Open Science Framework at: https://osf.io/jhfy9.

Funding: This study was funded by the Europäischer Fonds für regionale Entwicklung

(EFRE-0801767) and Bundesministerium für Bildung und Forschung (01KC2308A). ABB is funded by the Volkswagen Stiftung 'Impuls für Wissenschaft' grant. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Breast cancer is the leading cause of cancer-related death among women worldwide, responsible for approximately 66,600 deaths globally in 2022 [1,2]. Among its subtypes, metastatic breast cancer (MBC) represents the most severe form [3]. Triple-negative breast cancer (TNBC), characterized by the absence of estrogen, progesterone, and HER2 receptors, has an aggressive clinical course, with earlier age of onset, high recurrence rates, greater metastatic potential, and shorter overall survival compared to other breast cancer subtypes [4–6]. Standard approaches to prolong survival and improve the quality of patient life include chemotherapy, radiotherapy, hormone therapy, targeted therapy, immunotherapy, and surgery [6,7]. However, MBC's genetic heterogeneity and variable clinical behavior make it challenging to identify effective treatments [3,5,8–10]. It is well known that the immune system plays a critical role in several fundamental biological processes including the removal of cancer cells expressing tumor-specific antigens and the control of metastases. Nevertheless, cancer cells can evade immune surveillance through various mechanisms, some of which remain unidentified [11–14]. Building on that, tumor immunotherapy and targeted therapy have become therapeutic strategies to stimulate antitumor immune responses [15,16]. Despite advances, developing therapies to effectively control cancer and reduce metastasis remains an unmet need in cancer research [6,17,18]. On one hand, cancer is a complex, polygenic disease influenced by multiple environmental factors. On the other hand, discrepancies between animal models and human cancers, as well as variations in preclinical study designs impact translatability [19,20]. While new animal models can more accurately mimic human cancers by precise control over mutations [21], yet, choosing an optimal model to resemble the manifestations of MBC in humans remains challenging. Murine models particularly aid in basic research for understanding tumor biology and evaluating checkpoint inhibitors or other therapies [22–25]. Xenograft breast cancer models play a crucial role in preclinical trials, assessing drug efficacy and providing outcomes relevant to human breast cancer [19,26]. These models are used to evaluate primary tumor growth, metastatic activity, and response to therapy scenarios in both spontaneous and experimental settings [27]. Breast cancer tumors can be established in preclinical studies by choosing specific cell lines and injection sites based on tropism [28,29]. For instance, orthotopic implantation of cancer cells in the mammary fat pad creates a disease-relevant tumor microenvironment (TME) that closely resembles human breast cancer stages for better translation into the clinic [29,30]. Conversely, in subcutaneous models, the tumor cells are engrafted under the skin on the flank of the animal. Since the setup and tumor growth measurements are easier in these models, it is widely used for preclinical immunotherapies that target specific immune cell populations. However, it fails to metastasize and cannot accurately imitate the tumor microenvironment or the patient's response in clinical studies [30–32]. Most of the complications associated with breast cancer are due to metastasis developing in regional lymph nodes or distant organs including bone, lung, liver, and brain [28,33,34]. Genetically modified and immunocompromised mouse models have long been used to investigate the progression and

underlying mechanisms of breast cancer metastasis [27,28]. In these models, human breast cancer cells can be injected via various routes (subcutaneously, intravenously, intracardially, or orthotopically) with outcomes depending on the route used [27,28,35]. While tail vein injections primarily lead to lung metastasis, portal vein injection provokes liver colonization, and intracardiac infusions target multiple organs, including bone [29,36]. Over the past decade, immunotherapy has shown promise in breast cancer animal models, yet questions remain regarding the reproducibility and predictability of these models [24,25]. Thus, it is essential to examine the effects of immunotherapy in relevant preclinical models of MBC and explore sources of heterogeneity. This study aimed to systematically review published preclinical animal models of metastatic breast cancer (MBC) to evaluate the impact of immunotherapy on both primary and metastatic tumor regression compared to healthy controls through meta-analysis, as well as to investigate key study design variables essential for reproducibility in future preclinical studies.

Materials and methods

This systematic review was conducted according to a preregistered protocol in PROSPERO (CRD42021207033; available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=207033) to address the research question 'Does immunotherapy have successful effect on primary and metastatic tumor regression in the treated animal models with metastatic breast cancer, compared to control?'. The question was structured according to the PICO format with P as an animal model of metastatic breast cancer (all species), I as immunotherapy, C as untreated controls or vehicle only, and O as regression of primary and metastatic tumors (volume, area, weight). Neither risk of bias nor quality assessment were pre-planned. However, we reached a decision to assess the reporting quality based on the ARRIVE 2.0 guidelines [37,38]. There were no other deviations from the intended methodology as outlined in the protocol. The tumor model, induction method, cell line application route, drug administration route, and animal strain were analyzed as potential sources of heterogeneity.

Search strategy

The systematic search and removal of duplicates were conducted by one author (YM). The databases MEDLINE via PubMed, Embase via embase.com, and Web of Science Core Collection via Webofscience.com RWTH Aachen university were searched to identify all potentially relevant publications until the end of February 2024. We did not examine references of included studies for additional sources. The full search strategy was based on the search components: treatment or intervention using immunotherapeutic agent, metastatic breast cancer, animal model, and tumor regression. We combined keywords and MeSH terms for PubMed and adapted this strategy to the other databases (the full search strategies are available at S1 File). The filter of "Other animals" in the PubMed database, the animal filter in Embase, and Web of Science animal related topic fields were added to the search to remove non-animal studies. Since immunotherapy is a new treatment approach for MBC, to avoid too many irrelevant basic studies, a publication date filter (Published 2010 - Feb 2024) was also applied. The returned records were transferred to EndNote version X9.3.3 (RRID: SCR_014001) to check for overlap and deduplication of our library. The final results file was then transferred to Rayyan QCRI (RRID: SCR_017584, available at http://rayyan.qcri.org) to perform the initial screening of titles and abstracts by two independent reviewers (YM and MH) with blinding of the selection/decision process. Here, we checked again the searched items for any remaining duplicates.

Inclusion and exclusion criteria

Original controlled animal studies published in the English language using immunotherapeutic agents to achieve the main goal of treatment of metastatic breast cancer disease were included. No restrictions in intervention administration were used. Irrelevant studies were excluded based on predefined criteria. Exclusions entailed opinion articles, observational studies, case reports, reviews, clinical studies, *in vitro* studies, and treatment with nutrients or natural products. All

experimental animal species of any age and sex were included in our review. Excluded populations were humans, wildlife, domestic animals, zoo animals, endangered species, and exotic animals (S2 Table). Full-text screening of the potentially eligible articles for final inclusion was subsequently carried out by the two independent reviewers (YM, MH) and articles were included if they met our pre-specified inclusion criteria. The assessment of inter-rater reliability between the two reviewers was considered according to Cohen's kappa agreement and calculated using the Cohen's kappa calculator tool [39,40]. Here, a minimum 0.61 agreement was required for inclusion [41]. Any disagreement over the eligibility of particular studies was resolved through discussion between the above-mentioned reviewers. A third reviewer (JS) was consulted in case the first two reviewers could not reach an agreement.

Data extraction

One reviewer (YM) assessed each included article to extract bibliographic data (first author, year of publication, journal), data on study design (randomization, method of breast cancer induction, tumor size at the beginning time of intervention), animal model characteristics (animal species, strain, age, sex, weight, and number), intervention characteristics (type of treatment and control agents, dosage regimen and route of administration), primary outcomes: tumor weight, tumor volume, tumor size, and secondary outcomes: relative organ weights, body weight changes. The data were recorded into a purpose-built Microsoft Excel spreadsheet (RRID: SCR_016137). We first tried to extract numerical data from tables, or text. In studies where numerical data were not reported and presented only graphically in figures, a screenshot of the graphical data was taken, and adequate estimation of the outcome measurements was extracted from graphs using WebPlotDigitizer (RRID: SCR_013996). In studies where data were unobtainable, an initial request for unpublished information to the corresponding author was sent by email. If there was no response to our initial email after a minimum of seven business days, a second reminder email was sent. When there was still no response after a minimum of 30 business days following the initial email and data were missing, the results were excluded from the analyses. In cases where the outcome was measured repeatedly at different time points, only data from the endpoint (the last day of the experiment) was included. When more than one experiment was reported in the same manuscript, the experiments were only included as separate comparisons in the analyses if separate animals were used as the treatment group. If one control group was used for multiple intervention groups, the number of animals in the control group was divided by the number of intervention groups for meta-analysis. Extracted data were discussed in case of uncertainty of eligibility and final decisions were taken in a meeting with a second reviewer (JS).

Reporting quality assessment

The reporting quality was assessed for each of the included studies based on the Animal Research: Reporting of *In Vivo* Experiments reporting guidelines (ARRIVE 2.0). Each study was assessed against the following reporting items: study design, sample size, inclusion and exclusion criteria, animal randomization to different groups, blinding, animal characteristics, experimental procedure, statistical methods used, outcome measurements, and results. To assess reporting quality, for each section mentioned in the ARRIVE Essential 10 guidelines, studies were scored as "Fully reported" if all the information for each criterion was reported, "Partially reported" if only some of the information for each criterion was reported, and "Not reported" if none of the information for each criterion was reported. Reporting quality was expressed as a summary plot using Risk-of-bias Visualization (*robvis*) tool [42].

Data synthesis and statistical analysis

We considered meta-analyses feasible if homogeneity in the interventions, comparisons, and outcomes was present. In this manner, we divided the studies into comparisons that evaluated primary tumor regression and those that measured metastatic tumor regression. Meta-analysis was performed using R V.4.0.3 (RRID: SCR_009175) if a minimum of 4 comparisons per each primary and metastatic group were obtained. Since in all comparisons, we had a mean outcome score

and a measure of its variance, we calculated a normalized difference in means, or a standardized difference in means. For primary tumor volume and tumor weight datasets, as data existed on a ratio scale and sham data could be inferred, we presented normalized mean difference [39]. This value tells us the direction and the magnitude of the treatment effect with the effect sizes typically falling between −100% and + 100%. For metastatic tumor nodules number and tumor area, we presented continuous data with the standardized mean difference (SMD) with a 95% confidence interval (CI). In studies where standard error of the mean (SEM) was provided, SD was calculated by multiplying the square root of the number of animals used in the study by the SEM value ($s.d. = s.e. \times \sqrt{n}$). The random effects model was assigned to assess weights and the variation among studies. Additionally, we calculated the 95% prediction interval (PI) to have an estimated range that the true effect of a new study from the population of studies will fall in 95% of the cases. To have unbiased estimates of variance and covariance parameters, variance estimates were calculated using the restricted maximum likelihood (REML) approach and the degree of heterogeneity was quantified using the I^2 statistic index. Here, based on I^2 estimates in relation to the cumulative number of comparisons and their 95% CIs, we considered I^2 of 50% to 70% as representing moderate heterogeneity, and I^2 greater than 70% as a considerable heterogeneity. We performed uni- and multivariable meta-regression to explore the model-related possible sources of heterogeneity including the tumor induction method, cell line application site, animal strain, animal sex (where reported), and route of drug administration. Heterogeneity was described using Q (heterogeneity statistic), tau^2 (estimation of between-study variance), residual I^2 (the percentage of the residual variation that is attributable to between-study heterogeneity), and adjusted R^2 (the proportion of between-study variance explained by the covariate). Plots were created using the forest or orchard 2.0 R package [43]. Visual inspection of funnel plots, followed by Egger's regression test, a weighted linear regression of the treatment effect on its standard error, were used to statistically test for asymmetry arising from small study effects and possible publication bias. The 'trim and fill' method [44,45] was used to correct the pooled effect estimate for potential funnel plot asymmetry.

Results

Study selection and search results

The initial search yielded in total 3306 records, of which 2227 records were from PubMed and a total of 650 articles from Embase and 429 articles from Web of Science (Fig 1). After duplicates were removed (n = 1051), there were 2255 unique studies, which were assessed for eligibility at title and abstract stage. 368 studies were included at title and abstract stage and underwent full-text analysis to judge whether they met the predefined criteria. Ultimately, 108 articles fulfilled our inclusion criteria and were included for qualitative analysis. Here, the calculated agreement based on Cohen's kappa guidelines was 0.75 with percentage of agreement 88.88%, demonstrating a substantial agreement. Where publications included experiments on more than one model or with different treatment strategies, these were counted separately as individual comparisons. Three of these studies couldn't be included in the quantitative meta-analysis due to incomplete data such as the animal number per group. The relevant data were extracted from the remaining 105 articles containing 293 comparisons and used for meta-analysis. In one paper (S11 File, reference number 65), we recognized a discrepancy between the outcome data mentioned in the results and the values showed in the figure. Here, based on the experiment and the likelihood of text accuracy, we decided to include outcome data mentioned in text, in our meta-analysis.

Study characteristics

The included studies were published between 2010 and the end of February 2024 (S3 Table). All studies investigated mouse models and had ethical approval. In the majority of studies, female animals (97.2%) were used, and 3 studies reported using male animals. Out of 108 studies, 105 reported animal number per group either in the 'methods' or 'results' section but no studies discussed how the sample size was calculated. The age of the animals was reported in 91 studies (84.2%) and 15.7% of the studies reported the weight of the animals (S4 Table). In general, different methods for tumor induction were used (syngeneic, xenograft, Genetically Engineered Mouse (GEM), and patient-derived xenograft (PDX))

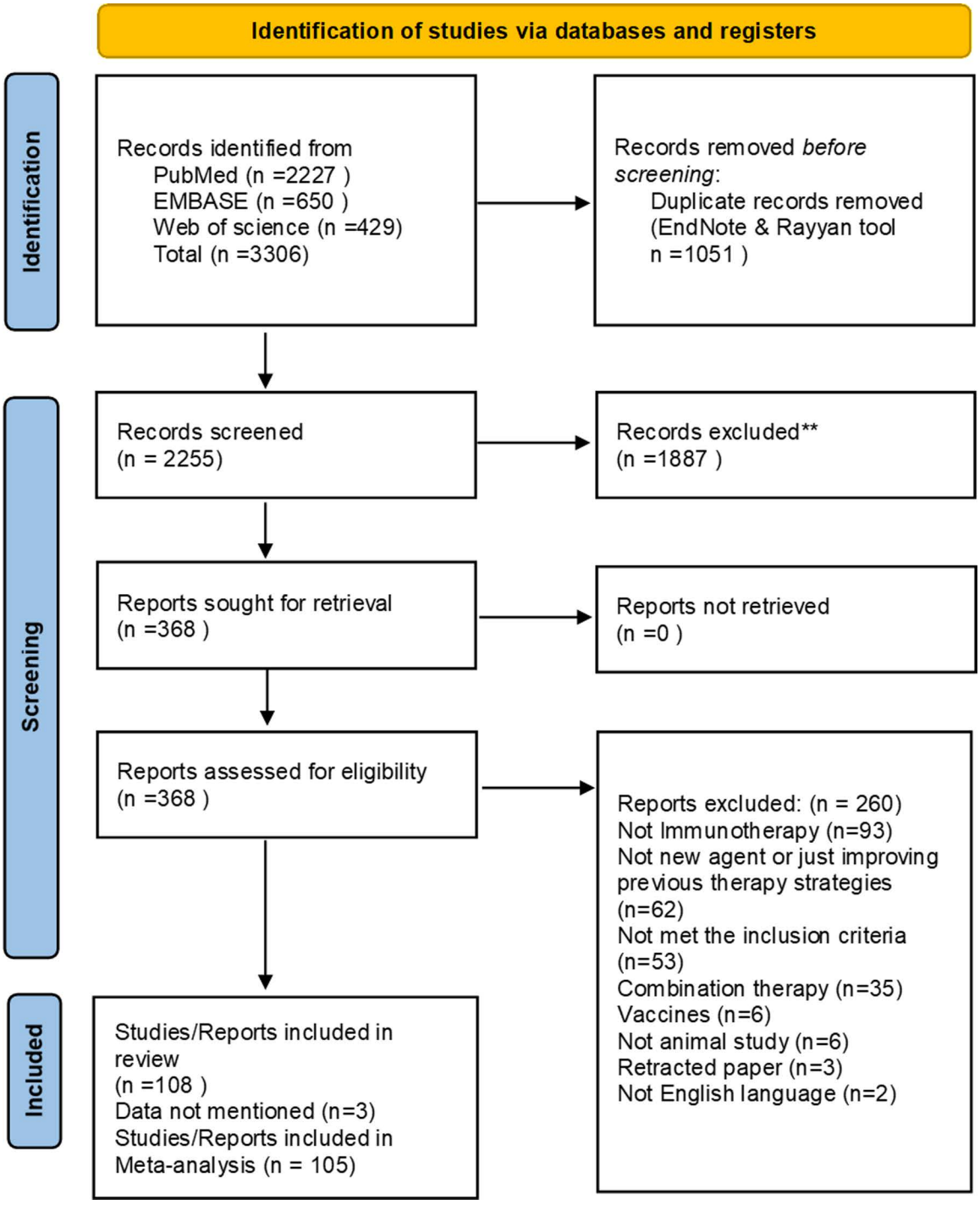


Fig 1. PRISMA flow chart of the study selection.

https://doi.org/10.1371/journal.pone.0322876.g001

and animal strain and inoculation site were reported accordingly in all studies. However, 8 papers that used cell lines to induce tumors didn't mention the injected cell number (S5 Table). Regarding the therapeutic interventions used, all studies reported the route of agent administration, and 98.1% also reported the dosage. Study duration was mentioned in all

studies and varied from 14 to 105 days. However, only 103 studies reported the starting time of drug intervention either according to tumor volume at the indicated day or the day number after inoculation (S6 Table).

Reporting quality assessment of the included studies

Of the 108 studies included in this SR, 73 studies (67.5%) reported randomization to control or treatment groups. However, none of these studies mentioned the method used for randomization. Moreover, measures to reduce performance bias (random housing and blinding of animal caretakers) were not reported nor was it stated whether outcome assessors were aware of the group allocation at any stage of the experiment or data analysis. Study design, outcome measures, statistical methods, and results were reported in all included studies, whereas none of them reported the method by which the sample size was chosen. Additionally, the criteria used for including and excluding animals, experimental units, or data points were partially provided. Only 14% and 43% of studies fully reported the experimental animal characteristics and experimental procedures accordingly (S7 Table). Visualization of the reporting quality scoring based on each criterion addressed in the ARRIVE Essential 10 guideline is shown in Fig 2.

Outcome measures and exploration of heterogeneity

Primary tumor regression and metastasis tumor regression were two specific types of outcomes in the included studies and were classified and assessed separately.

1. **Effect of immunotherapy on primary tumor regression.** A total of 233 individual comparisons from 95 studies assessed primary tumor regression, measured either by tumor volume (n = 214) or tumor weight (n = 19). As shown in Fig 3, the pooled effect size was statistically significant, favoring immunotherapeutic agents to reduce both primary tumor volume and weight by over 50%. Specifically, there was a reduction of 58.73% (NMD, 95% CIs [55.78; 61.68], $P < 0.0001$, n = 214) for primary tumor volume, and 52.78% (NMD, 95% CIs [38.08; 67.47], $P < 0.0001$, n = 19) for primary tumor weight. However, a random-effects model showed considerable heterogeneity across the included studies for both outcomes. For tumor volume, heterogeneity was $I^2 = 98.05\%$, $\tau^2 = 10294.25$, $P < 0.0001$(n = 214), and for tumor weight, heterogeneity was $I^2 = 98.8\%$, $\tau^2 = 1043.87$, $P < 0.0001$ (n = 19). Prediction intervals further illustrated the high level of heterogeneity with a wide range positive effect sizes for primary tumor volume [19.87; 97.79]. For tumor weight, the prediction interval crossed the line of no effect, even including negative effect sizes [-12.22; 117.79]. In the tumor volume dataset (n = 214), comprising 211 comparisons in female mice and 3 in male mice, 5 experiments used genetically engineered mouse (GEM) models, 11 used patient-derived xenografts (PDX), 30 used syngeneic models, and 169 used xenografts. The most frequently used cell lines for tumor induction were MDA-MB-231 (n = 94), 4T1 (n = 22), and MCF7

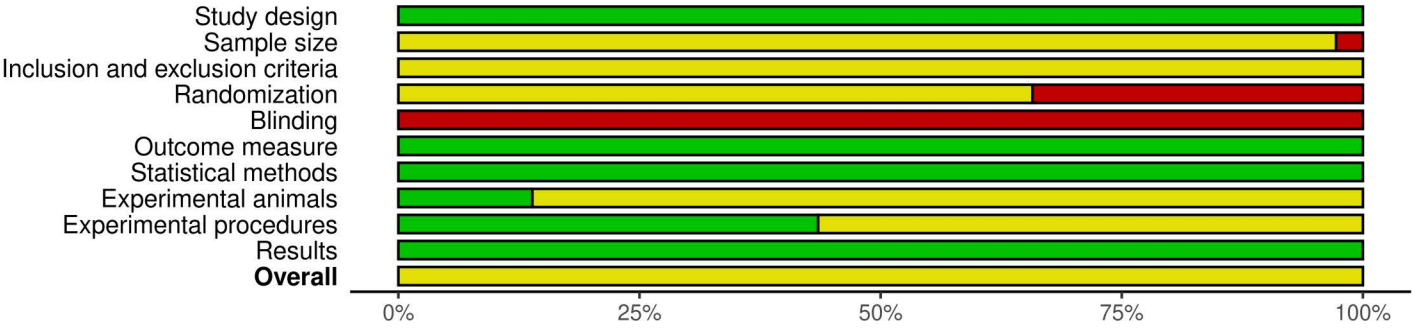


Fig 2. Summary plot of reporting quality assessment based on ARRIVE Essential 10. Green indicates fully reported items, Yellow indicates partially reported items, and Red indicates items not reported (generated using the *robvis* tool). Overall, the studies only partially reported the information required by the ARRIVE guidelines.

https://doi.org/10.1371/journal.pone.0322876.g002

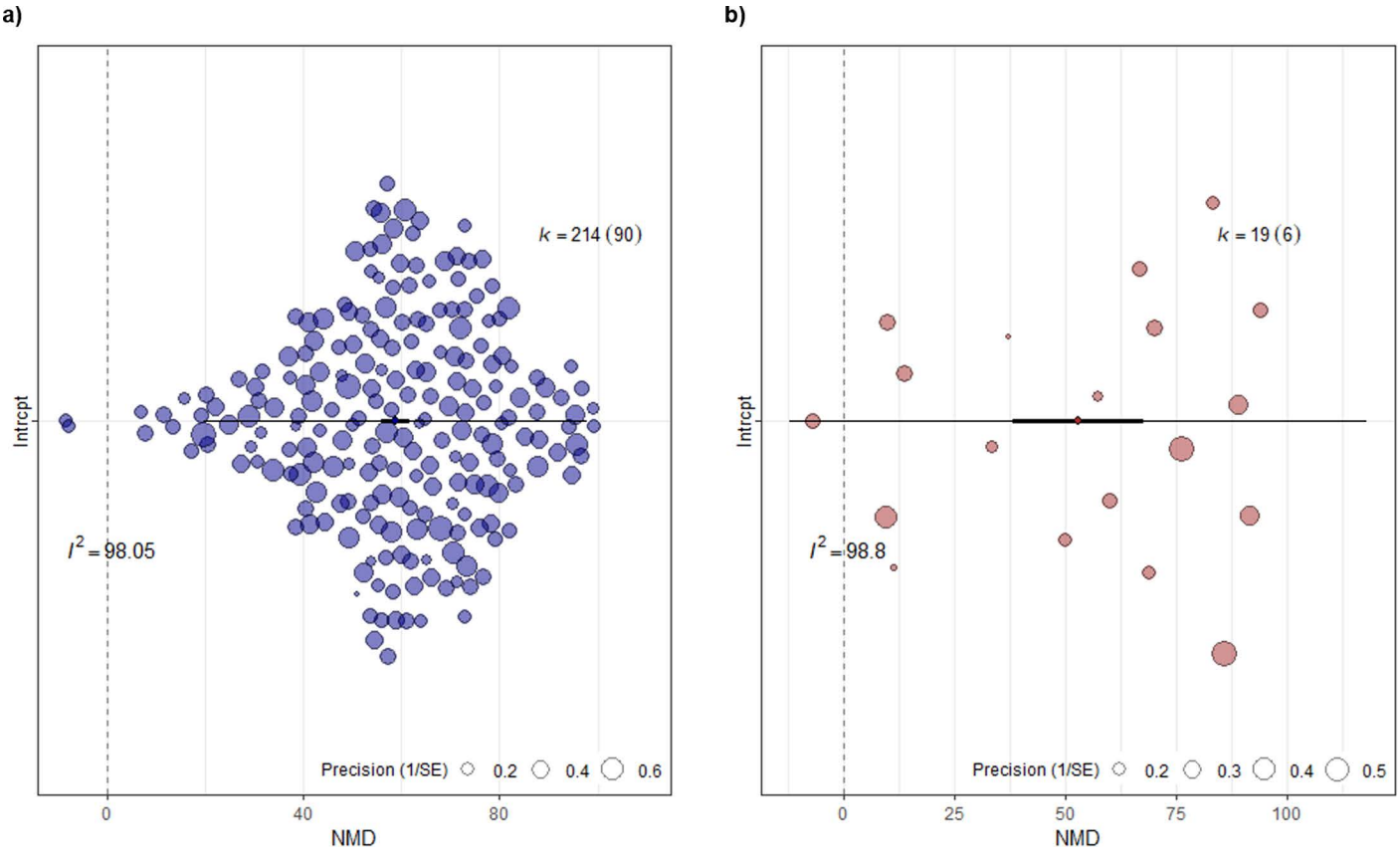


Fig 3. Orchard plots of meta-analysis results for primary tumor regression. a) Tumor volume, and b) Tumor weight. On each plot, the thick black line indicates 95% confidence intervals (CI), the extended black line indicates 95% prediction intervals (PI), and the midpoint is denoted by a diamond indicating the pooled normalized mean difference (NMD) effect estimate. Each circle represents a comparison and its diameter denotes the weight each comparison bears in the pooled effect size based on precision, meaning the larger the sample size the comparison has, the more precise it is and the higher its weight in the pooled effect. I^2 reports the quantitative value for heterogeneity (98.05% and 98.8% respectively). The number of comparisons included in the meta-analysis is denoted by *k*, and with the number of unique studies denoted in brackets.

https://doi.org/10.1371/journal.pone.0322876.g003

(n = 22). For MDA-MB-231 and MCF7 cell lines, subcutaneous injections were more common, whereas the 4T1 cell line was more frequently injected in the mammary fat pad. In general, female *BALB/c nude* (n = 137) and *BALB/c* (n = 24) mice were the major used strains as immunosuppressant and immunocompetent mice models, respectively. In contrast, among the 19 comparisons that used primary tumor weight as the outcome, most (n = 11) used female *BALB/c* mice with 4T1 cell line injected in the mammary fat pad as an orthotopic-syngeneic model. Other cell lines used included MX1, MDA-MB-231, and MCF7.

1.1. Exploring heterogeneity in the effects of immunotherapy on tumor volume and weight: Uni- and multivariable meta-regression analyses were conducted to examine potential sources of heterogeneity in the effects of immunotherapy on tumor volume and weight. We explored several biological and methodological factors as predictors. Specifically, we assessed whether tumor model type (syngeneic, xenograft, GEM, PDX), cell lines, cell line application route, animal sex, animal strain, and immunotherapeutic agent administration route could explain any of the observed between-study variation. In the univariable meta-regression for tumor volume outcomes, no association was found between the selected predictors and effect size, as the p-value for the test of moderators was greater than 0.05 for all predictors (Fig 4). The residual heterogeneity, or unexplained variance, remained significant, as shown by high I^2 values

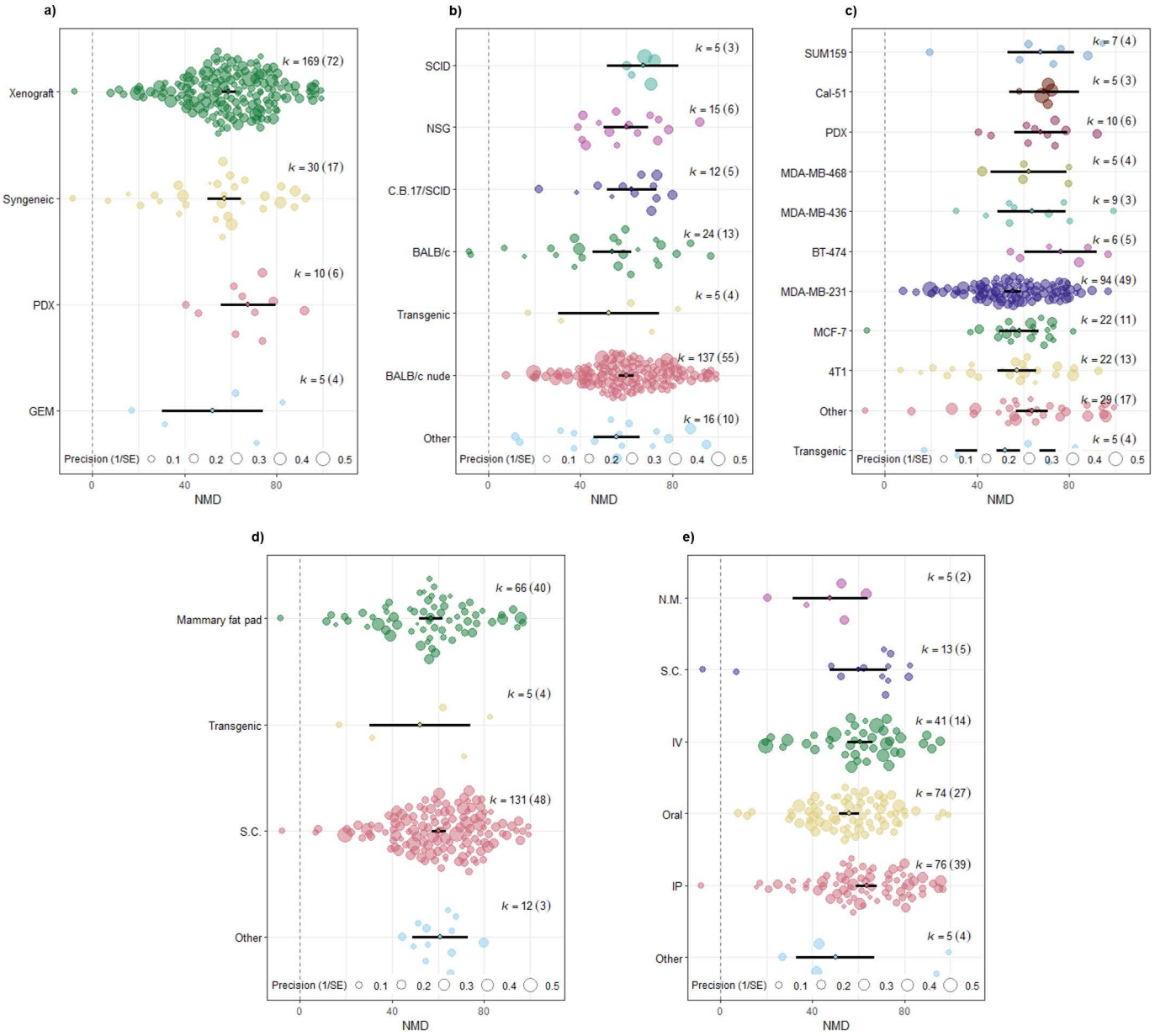


Fig 4. Orchard plots of univariate meta-regression analysis for primary tumor volume. a) Tumor model, b) Animal strain, c) Induction method, d) Cell line application route, and e) Drug administration route. The thick black line indicates 95% confidence intervals (CI). Each circle represents a comparison and its diameter denotes the weight each comparison bears in the pooled effect size based on precision. No difference was observed between subgroups and the p-value for test of moderators was not statistically significant (P = 0.45, P = 0.67, P = 0.10, P = 0.63, P = 0.12, respectively). The number of comparisons included in each level of the variable of interest is denoted by *k*, and with the number of unique studies denoted in brackets.

https://doi.org/10.1371/journal.pone.0322876.g004

across predictors: animal strain ($I^2 = 87.07\%$), tumor model ($I^2 = 87.28\%$), induction method ($I^2 = 86.13\%$), cell line application route ($I^2 = 87.35\%$), and drug administration route ($I^2 = 86.39\%$), all with $P < 0.0001$. Univariable meta-regression based on animal sex was not conducted due to only three comparisons involving male animals, which would render the results unreliable. Additionally, we conducted a multivariable meta-regression using a mixed-effects model on the tumor volume dataset (214 comparisons), using the above predictors as moderators (S8 Table). This analysis revealed that the set of

moderators (coefficients 2–24) significantly explained variation in effect sizes (test of moderators: F(df1 = 23, df2 = 190) = 1.6541, P = 0.0363). However, the amount of variance explained by these moderators was relatively low, as indicated by the test for residual heterogeneity (QE(df = 190) = 1375.8701, P < .0001, R^2 = 9.10%). This indicated that the moderators accounted for only a small fraction of the total variance in outcomes, suggesting that additional unmeasured factors may contribute to the observed differences between studies. Due to the limited number of unique studies (n = 6) with only 19 comparisons, meta-regression analysis was not conducted for the tumor weight dataset.

2. Effect of immunotherapy on metastatic tumor regression. Thirty-two studies evaluated metastatic tumor regression, and from these, 60 individual comparisons were included in the meta-analysis. Of these comparisons, 51 assessed tumor regression based on a reduction in the number of metastatic nodules in the lung, while 9 used tumor area reduction measured by bioluminescence. To provide clarity, the overall effect estimate was calculated separately for each dataset. For studies counting lung nodules, the overall effect of immunotherapeutic agents was significant (2.63 SD, 95% CIs [1.89; 3.37], P < 0.0001, n = 51), as was the effect in studies using bioluminescence (1.79 SD, 95% CIs [1.06; 2.53], P < 0.0001, n = 9), indicating a positive impact of these agents on metastasis regression. However, as shown in Fig 5, the pooled estimates indicated considerable between-study heterogeneity in both datasets. Specifically, for the lung

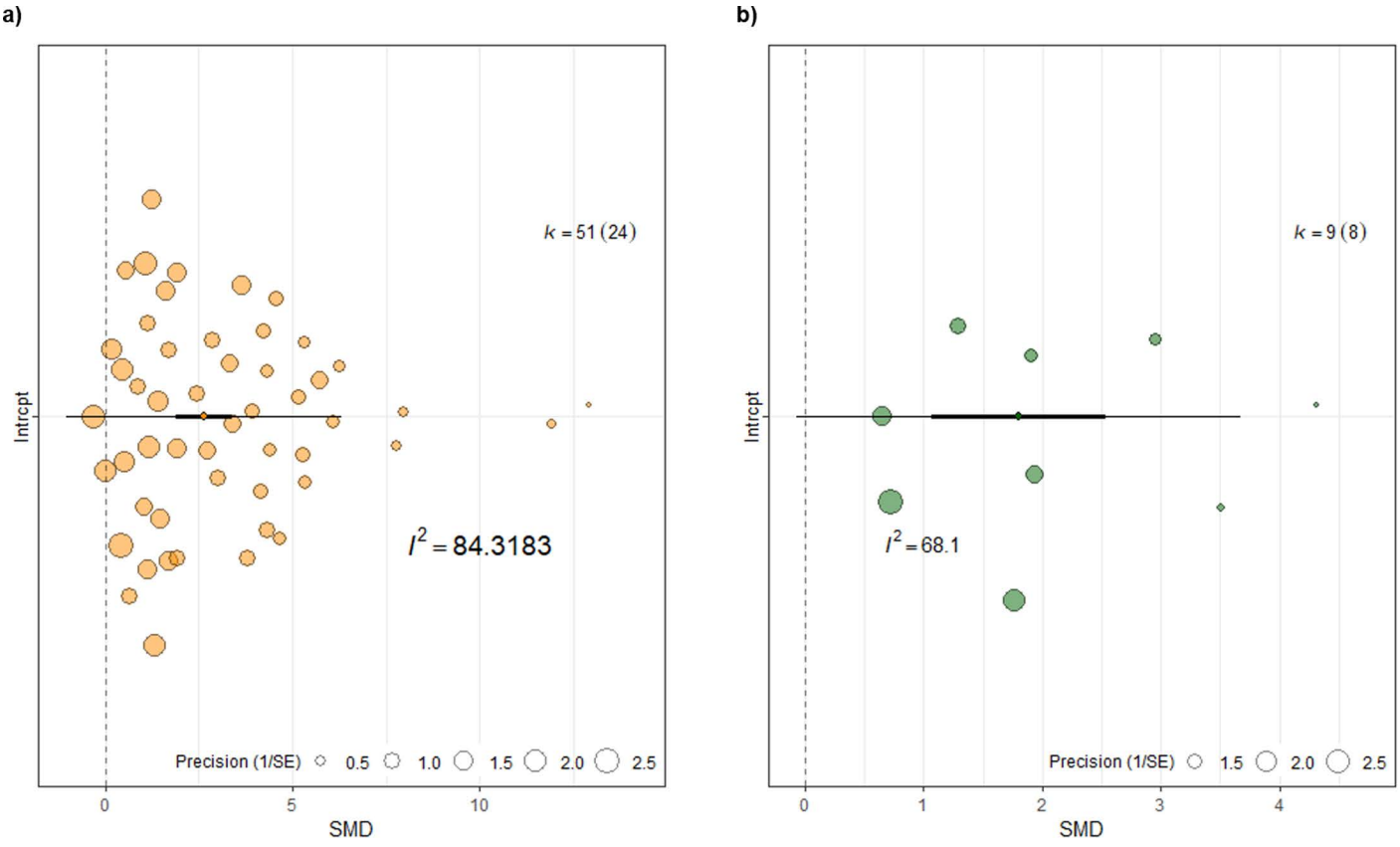


Fig 5. Orchard plots of metastatic tumor regression. a) Metastatic nodules number, and b) Metastatic tumor area based on bioluminescence measurement. On both plots, the thick black line indicates 95% confidence intervals (CI), the extended black line indicates 95% prediction intervals (PI), and the midpoint denoted by a diamond indicating the pooled standardized mean difference (SMD). Each circle represents a comparison and the size of it denotes the weight each comparison bears in the pooled effect size based on precision meaning the bigger sample size the comparison has, the more precise it is. I^2 reports the quantitative value for heterogeneity (84.32% and 68.1% respectively). The number of comparisons included in the meta-analysis is denoted by *k*, and with the number of unique studies denoted in brackets.

https://doi.org/10.1371/journal.pone.0322876.g005

nodule dataset, heterogeneity was Q(df = 50) = 240.8317, I^2 = 84.32%, P < 0.0001 (n = 51), and for the bioluminescence dataset Q(df = 8) = 24.8111, I^2 = 68.09%, P = 0.0017 (n = 9). The prediction intervals further highlighted this heterogeneity and spanned the line of no effect, indicating potential negative effect sizes (PIs [-1.03, 6.31] for lung nodules and PIs [-0.07, 3.66] for bioluminescence). Among the 51 comparisons counting metastatic nodules, different tumor models were employed: 3 used GEM, 1 used PDX, 27 used syngeneic, and 20 used xenograft models, all with female mice. The most commonly used cell lines were 4T1 (n = 20) and MDA-MB-231 (n = 18). For the 4T1 cell line, all studies used female *BALB/c* mice, with 17 comparisons using the mammary fat pad as the injection site, 2 using intravenous, and 1 using subcutaneous injection. On the contrary, for the MDA-MB-231 cell line, the most used route was intravenous (n = 9). Overall, female *BALB/c* and *NSG* mice were the most used strains as immunocompetent and immunosuppressant mice models, respectively. In the 9 comparisons assessing tumor area reduction by bioluminescence, the xenograft model using the MDA-MB-231-Luc cell line injected intravenously was the major tumor induction method (n = 6). Only three comparisons used the syngeneic model with the 4T1-luc cell line injected into the mammary fat pad.

2.1. Exploring heterogeneity in the effect of immunotherapy on metastatic tumor regression: We conducted univariable and multivariable meta-regression analyses to investigate potential sources of heterogeneity in the metastatic nodule reduction dataset. In the univariable meta-regression analyses, a considerable percentage of residual heterogeneity remained (Fig 6), and none of the selected predictors could explain the observed between-study differences. Specifically, residual heterogeneity values were as follows: animal strain (I^2 = 87.86%), tumor model (I^2 = 87.11%), induction method (I^2 = 86.67%), cell line application route (I^2 = 88.16%), and drug administration route (I^2 = 85.18%), with P < 0.0001 for all. All experiments in this dataset used female animals, so animal sex was not considered as a variable. The multivariable meta-regression also failed to account for the heterogeneity observed in the dataset. The test for residual heterogeneity remained significant QE(df = 35) = 132.7078, P < .0001, and the test of moderators (coefficients 2–16) did not reach statistical significance (F(df1 = 15, df2 = 35) = 1.3944, P = 0.2040; see S9 Table). Given the limited number of unique studies (n = 8), we decided not to not to statistically investigate the study design characteristics as sources of heterogeneity for metastatic tumor area reduction using bioluminescence measurements.

Assessment of publication bias

Visual inspection of the funnel plots suggested asymmetry among the larger studies for primary tumor volume. Additionally, asymmetry was observed in the funnel plot of the metastatic tumor nodules dataset, indicating the potential presence of small study effects and suggesting that some studies might be missing from our dataset, potentially biasing our findings (S10 Fig). For the primary tumor volume dataset, Egger's regression test did not indicate significant funnel plot asymmetry (t = -0.5017, df = 212, p-val = 0.6164). However, a trim and fill analysis imputed 32 theoretically missing studies (SE = 9.5271), resulting in an adjusted effect size of 54.35% (P < 0.0001), compared to the unadjusted effect size of 58.73% NMD (Fig 7a). For the metastatic tumor nodules dataset, both the trim-and-fill analysis and Egger's regression indicated funnel plot asymmetry and potential publication bias. The intercept of the regression model was significantly greater than zero, confirming asymmetry in the funnel plot (Egger's test: intercept = 2.5291, t = 3.959, P = 0.000243). The trim-and-fill method further supported this by identifying statistically significant funnel plot asymmetry and suggesting the presence of potentially missing studies, which represent less effective immunotherapy experiments with lower effect sizes. The analysis estimated 13 missing studies (SE = 4.6794), resulting in an adjusted pooled effect size of 1.88 SD (P < 0.0001) compared to 2.63 SD in the original random-effects meta-analysis (Fig 7b). Although the trim-and-fill analysis adjusted the estimated effect size for both datasets, favoring immunotherapy over control, significant heterogeneity remained between studies (primary tumor volume: tau^2 = 392.7627, SE = 46.0535, I^2 = 90.34%, P < 0.0001, n = 246; metastatic nodules number: tau^2 = 6.0356, SE = 1.3076, I^2 = 89.22%, P < 0.0001, n = 64). Therefore, the adjusted intervention effect was not considered a more valid estimate of the overall effect of immunotherapy. The results should be interpreted with caution, as the inclusion of imputed missing studies likely extended the range of effect sizes and increased

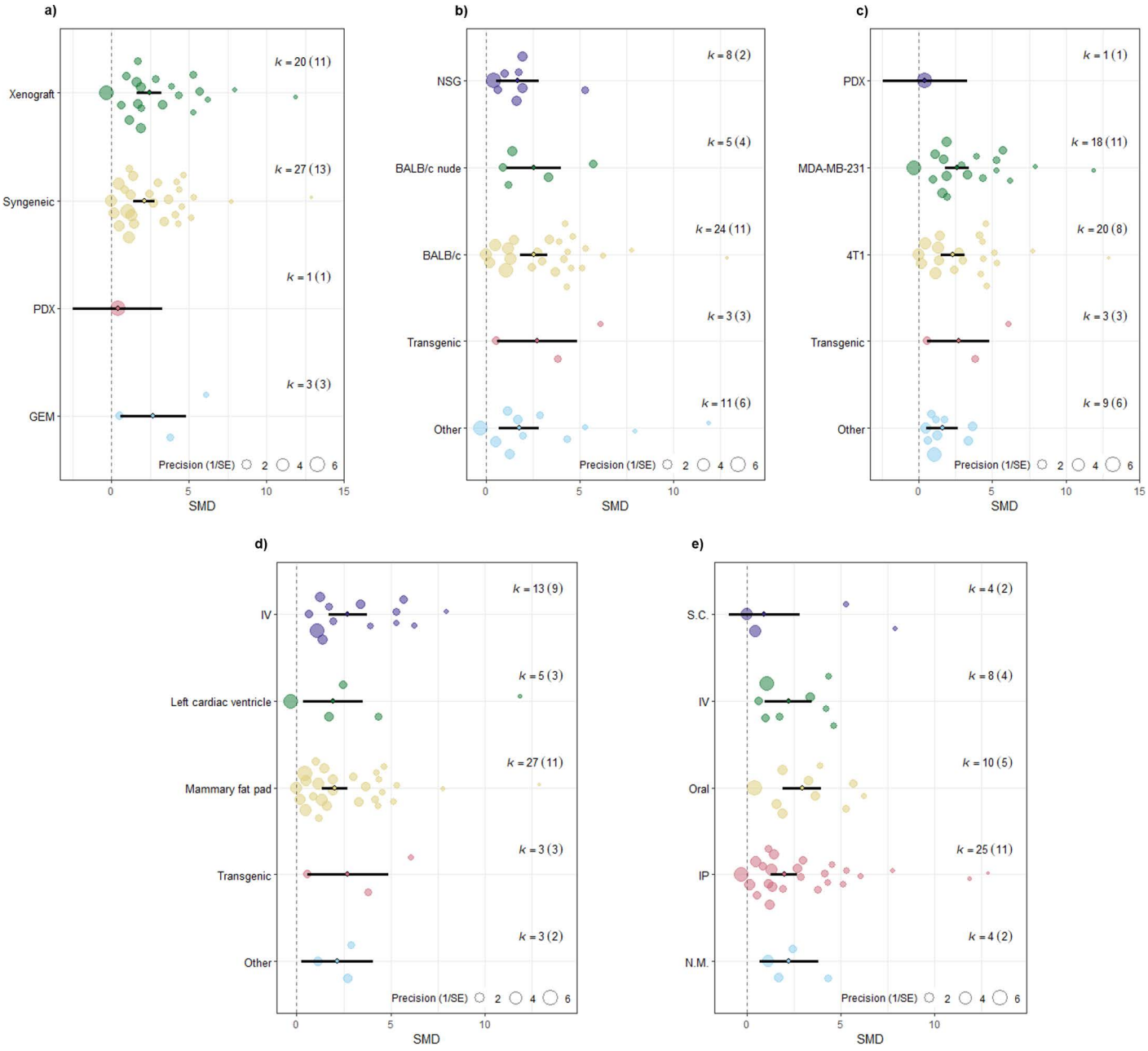


Fig 6. Orchard plots of univariate meta-regression analysis for metastatic nodules number. a) Tumor model, b) Animal strain, c) Induction method, d) Cell line application route, and e) Drug administration route. The thick black line indicates 95% confidence intervals (CI). Each circle represents a comparison and its diameter denotes the weight each comparison bears in the pooled effect size based on precision. No difference was observed between subgroups and the p-value for test of moderators was not statistically significant (P = 0.42, P = 0.59, P = 0.55, P = 0.81, P = 0.39, respectively). The number of comparisons included in each level of the variable of interest is denoted by *k*, and with the number of unique studies denoted in brackets.

https://doi.org/10.1371/journal.pone.0322876.g006

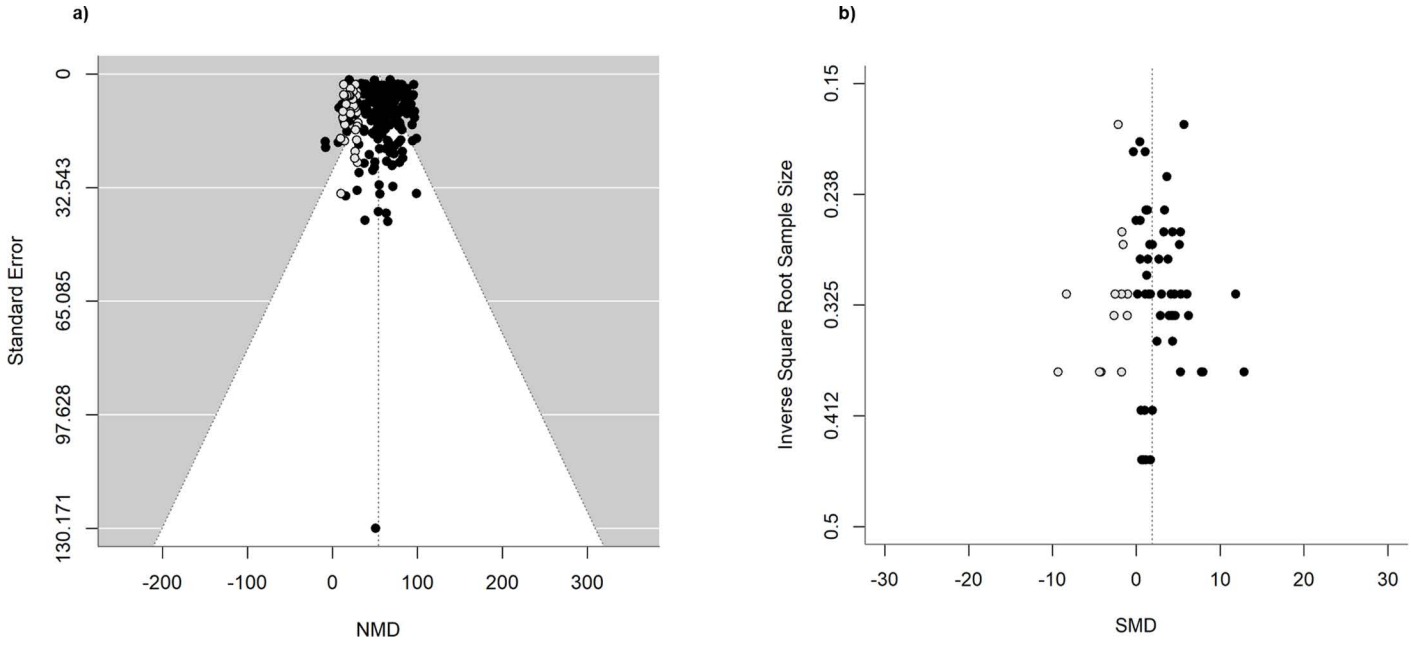


Fig 7. Funnel plots and Trim and Fill analysis of immunotherapy outcomes. a) Primary tumor volume, and b) Metastasis nodules number**.** The closed dots indicate the observed studies. The open dots indicate the missing studies imputed by the trim and fill method (n = 32 and n = 13 for a and b, respectively). The x-axis shows the observed outcome as normalized mean difference or the standardized mean difference with the y-axis representing the standard error or the inverse square root of the sample size, respectively. The two dotted lines on either side of plot (a) represent the 95% confidence intervals. The short-dash vertical line indicates the adjusted effect size including potentially missing studies under the random effects model: (a) NMD = 54.35%, and (b) SMD = 1.88 SD.

https://doi.org/10.1371/journal.pone.0322876.g007

heterogeneity among studies. Funnel plot analysis was not conducted for the primary tumor weight or metastatic tumor area reduction datasets due to the inclusion of fewer than 10 unique studies, which would result in low statistical power to reliably detect asymmetry.

Discussion

Breast cancer is a heterogeneous disease with variations in aggressiveness, metastasis potential, treatment response, and clinical outcomes [1,46]. While breast cancer's causes, nature, and risk factors have been extensively studied in humans [47–53], animal models remain essential for studying tumorigenesis, testing therapeutic strategies, and monitoring drug responses [54–57]. Previous literature has evaluated a range of animal models for breast cancer, pinpointing their unique features, strengths, weaknesses, and the principles of tumor model selection to fit the purpose of the investigation [20,25,58–61]. The herein study reviewed systematically the effects of immunotherapy in published animal models of MBC and assessed the quality of reporting in these studies. Additionally, we conducted a quantitative analysis to identify potential sources of heterogeneity and explore their implications for reproducibility in future preclinical studies.

Preclinical systematic reviews and meta-analyses differ from those in the clinical field. The greater variability in experimental designs, animal models, dosing regimens, and outcome measures often introduces substantial heterogeneity [62,63]. The strength of preclinical systematic reviews lies in exploring this heterogeneity to identify factors related to study design or quality that influence outcomes. To our knowledge, this is the first systematic review and meta-analysis focused on animal models of immunotherapy in MBC.

Our assessment of reporting quality based on parameters adapted from the ARRIVE guidelines revealed significant concerns that may compromise internal validity and reproducibility. Most of the included studies did not report key

information on methods to reduce biases, such as allocation, concealment, randomization, and blinding of outcome assessors. In studies where randomization was mentioned, details were insufficient for proper assessment. Although lack of reporting does not imply lack of implementation, previous systematic reviews has shown that animal experiments omitting these details tend to overestimate treatment effects [64]. Furthermore, crucial information about animal model characteristics (e.g., age, weight), experimental design (e.g., cell line, cell number, administration route), and statistical methods (e.g., power calculations, sample size justification) was often missing. These findings are in line with previous studies that evaluate the internal validity of preclinical experiments [65–68]. Though we did not address external validity (e.g., representativeness of animal samples, clinical relevance of models, and animal–human species differences), it is essential to ensure internal validity and reproducibility to enhance the clinical translatability of preclinical findings [68,69]. Therefore, improving transparency and methodological rigor in reporting is vital for increasing the reproducibility and reliability of *in vivo* tumor models across research laboratories and achieving more translatable results for clinical trials.

A key finding of our study was that most immunotherapy studies in MBC models used xenografts in immunodeficient mice, particularly nude mice. Given their lack of functional T cells, these models are suboptimal for evaluating immune-based therapies, as they fail to replicate tumor–immune system interactions [70]. While xenograft models offer advantages in tumor engraftment and reproducibility, their limitations should be considered when interpreting immunotherapy outcomes [71]. Syngeneic and genetically engineered models with intact immune systems provide more physiologically relevant platforms for assessing immunotherapies [72].

For the quantitative analysis, studies were divided into primary and metastatic tumor regression categories to account for inter-tumoral heterogeneity and diversity between primary and metastasis tumors [73]. Studies measured tumor regression by weight or volume for primary tumors and by nodule count or area for metastatic tumors. Results indicated that immunotherapy had a significant, positive effect, with a reduction of over 50% in primary tumor volume and weight, as well as a reduction in metastatic tumor nodules number and area. However, statistical analyses revealed moderate to considerable heterogeneity, suggesting substantial variability in immunotherapy effectiveness across studies. Prediction intervals further showed high heterogeneity, with ranges that included negative effect sizes for primary tumor weight and both metastatic tumor datasets. This estimation is based on the effect sizes observed in the meta-analysis and broadly corresponds to their range. The heterogeneity likely reflects variability in study methodologies, such as randomization and selection processes, which may introduce potential selection bias. Consequently, the prediction interval should be interpreted as describing the range of observed effect sizes, rather than predicting the range of effect sizes expected in future studies [74].

We were particularly interested in whether characteristics related to the different tumor models used in the included studies might help to explain the observed heterogeneity. The diversity of tumor models in our review arises from the use of both transplanted and genetically engineered models. Additionally, factors such as the origin and type of tumor tissue or cells (animal or human), their quantity, and the site of implantation contribute to biological heterogeneity [25]. We examined model-related sources of heterogeneity through univariate and multivariable meta-regression analyses, but much of the variability remained unexplained, underscoring the need for further evaluation. Our data show that MDA-MB-231, MCF7, and 4T1 cell lines, typically administered via the mammary fat pad or subcutaneously, dominated primary tumor studies. In studies on metastatic tumor regression, the MDA-MB-231 and 4T1 cell lines were the most commonly used, with administration via the mammary fat pad or intravenous route.

Limitations

This systematic review and meta-analysis provided valuable insights into the effects of immunotherapy on MBC regression and an overview of the animal models used. However, our analysis has limitations. First, factors such as the injected cell number, cell environment (e.g., media or Matrigel®), tumor size at treatment onset, and the type and dosage of immunotherapeutic agent are known to influence tumor establishment and antitumor responses. However, limited data on these factors prevented a post-hoc analysis with a sufficient sample size to fully evaluate their impact on the unexplained

heterogeneity in our analyses. Second, small study effects and possible publication bias may influence our findings, as larger studies with significant results are more likely to be published [66,68,75]. The absence of neutral or negative results hinders the development of new models for drug testing, as it increases the risk of duplicating past failures. Additionally, it impedes efforts to identify optimal animal models, leading to unnecessary use of additional animals. Third, our analysis relied on the trim-and-fill method to assess publication bias, which assumes that funnel plot asymmetry is solely indicative of publication bias. While reliance on publicly available information, selective reporting of outcomes and analyses, and the absence of reports on unsuccessful interventions contribute to potential publication bias, observed heterogeneity and publication bias may also interact and influence each other. Publication bias can significantly impact heterogeneity estimates [76,77], and moderate to high heterogeneity can reduce the power of the trim-and-fill method [78,79]. As a result, our analysis was affected by both observed heterogeneity and potential publication bias. Moreover, the lack of uniformity and standardized protocols in preclinical studies makes it challenging to make generalized conclusions Therefore, findings from preclinical meta-analyses should be interpreted within this context.

Conclusions

Overall, this systematic review and meta-analysis highlighted the wide variety of tumor models used to investigate immunotherapeutic agents in MBC. Our findings underscored several factors that contribute to the translational gap and lack of reproducibility in these models. While our results showed that immunotherapy had significantly beneficial pooled effects on both primary and metastasis tumor regression in MBC, improved adherence to ARRIVE guidelines in reporting animal studies is needed. Moreover, our work encourages further systematic reviews on animal models used for other therapies across different cancer types. Collecting and organizing this data could support the development of an open-access database of animal models categorized by intervention and degree of tumor regression, allowing researchers to select the most suitable model for their study goals before initiating an animal experiment.

Supporting information

S1 File. Search strategy.
(DOCX)

S2 Table. Summary of study inclusion and exclusion criteria.
(DOCX)

S3 Table. Main bibliographic data of the 100 studies included in Systematic review.
(DOCX)

S4 Table. Main animal characteristics of the 100 studies included in Systematic review.
(DOCX)

S5 Table. Main study design characteristics of the 100 studies included in Systematic review.
(DOCX)

S6 Table. Main experiment design characteristics of the 100 studies included in Systematic review.
(DOCX)

S7 Table. Reporting quality assessment using the ARRIVE guidelines.
(DOCX)

S8 Table. Multivariable meta-regression analyses for primary tumor volume regression dataset.
(DOCX)

S9 Table. Multivariable meta-regression analyses for metastatic lung nodules number dataset.
(DOCX)

S10 Fig. Funnel plots of immunotherapy outcomes.
(DOCX)

S11 File. List of the included studies in Systematic review and Meta-analyses and PRISMA 2020 Checklist.
(DOCX)

S12 File. All studies identified in the literature search, including the reason(s) for exclusion.
(XLSX)

S13 File. All data extracted from the primary research sources for the systematic review and meta-analysis.
(XLSX)

Acknowledgments

We are grateful for the support of the CAMARADES Berlin team and wish to thank them for their valuable suggestions.

Author contributions

Conceptualization: Yalda Mirzaei, Alexandra Bannach-Brown, René H. Tolba, Julia Steitz.

Data curation: Yalda Mirzaei, Martina Hüffel.

Formal analysis: Yalda Mirzaei.

Funding acquisition: Julia Steitz.

Investigation: Yalda Mirzaei.

Methodology: Yalda Mirzaei, Sarah McCann, Alexandra Bannach-Brown.

Project administration: Yalda Mirzaei, Julia Steitz.

Resources: Yalda Mirzaei, Julia Steitz.

Software: Yalda Mirzaei.

Supervision: René H. Tolba, Julia Steitz.

Validation: Yalda Mirzaei, Sarah McCann, Alexandra Bannach-Brown, Julia Steitz.

Visualization: Yalda Mirzaei.

Writing – original draft: Yalda Mirzaei.

Writing – review & editing: Yalda Mirzaei, Martina Hüffel, Sarah McCann, Alexandra Bannach-Brown, René H. Tolba, Julia Steitz.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63. https://doi.org/10.3322/caac.21834 PMID: 38572751
2. Yi M, Li T, Niu M, Luo S, Chu Q, Wu K. Epidemiological trends of women's cancers from 1990 to 2019 at the global, regional, and national levels: a population-based study. Biomark Res. 2021;9(1):55. https://doi.org/10.1186/s40364-021-00310-y PMID: 34233747
3. Kennecke H, Yerushalmi R, Woods R, Cheang MCU, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. J Clin Oncol. 2010;28(20):3271–7. https://doi.org/10.1200/JCO.2009.25.9820 PMID: 20498394
4. Fan Y, He S. The characteristics of tumor microenvironment in triple negative breast cancer. Cancer Manag Res. 2022;14:1–17. https://doi.org/10.2147/CMAR.S316700 PMID: 35018117

5. Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol. 2016;13(11):674–90. https://doi.org/10.1038/nrclinonc.2016.66 PMID: 27184417
6. Sazgarnia A, Montazerabadi AR, Bahreyni-Toosi MH, Ahmadi A, Aledavood A. In vitro survival of MCF-7 breast cancer cells following combined treatment with ionizing radiation and mitoxantrone-mediated photodynamic therapy. Photodiagnosis Photodyn Ther. 2013;10(1):72–8. https://doi.org/10.1016/j.pdpdt.2012.06.001 PMID: 23465375
7. Liu B, Zhou H, Tan L, Siu KTH, Guan X-Y. Exploring treatment options in cancer: tumor treatment strategies. Signal Transduct Target Ther. 2024;9(1):175. https://doi.org/10.1038/s41392-024-01856-7 PMID: 39013849
8. Nielsen DL, Andersson M, Andersen JL, Kamby C. Antiangiogenic therapy for breast cancer. Breast Cancer Res. 2010;12(5):209. https://doi.org/10.1186/bcr2642 PMID: 21067536
9. Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. Molecular definition of breast tumor heterogeneity. Cancer Cell. 2007;11(3):259–73. https://doi.org/10.1016/j.ccr.2007.01.013 PMID: 17349583
10. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747–52. https://doi.org/10.1038/35021093 PMID: 10963602
11. Swann JB, Smyth MJ. Immune surveillance of tumors. J Clin Invest. 2007;117(5):1137–46. https://doi.org/10.1172/JCI31405 PMID: 17476343
12. Pinzon-Charry A, Maxwell T, López JA. Dendritic cell dysfunction in cancer: a mechanism for immunosuppression. Immunol Cell Biol. 2005;83(5):451–61. https://doi.org/10.1111/j.1440-1711.2005.01371.x PMID: 16174093
13. Pyfferoen L, Mestdagh P, Vergote K, De Cabooter N, Vandesompele J, Lambrecht BN, et al. Lung tumours reprogram pulmonary dendritic cell immunogenicity at the microRNA level. Int J Cancer. 2014;135(12):2868–77. https://doi.org/10.1002/ijc.28945 PMID: 24789737
14. Bates JP, Derakhshandeh R, Jones L, Webb TJ. Mechanisms of immune evasion in breast cancer. BMC Cancer. 2018;18(1):556. https://doi.org/10.1186/s12885-018-4441-3 PMID: 29751789
15. Wei J, Sun H, Zhang A, Wu X, Li Y, Liu J, et al. A novel AXL chimeric antigen receptor endows T cells with anti-tumor effects against triple negative breast cancers. Cell Immunol. 2018;331:49–58. https://doi.org/10.1016/j.cellimm.2018.05.004 PMID: 29935762
16. Tang H, Qiao J, Fu Y-X. Immunotherapy and tumor microenvironment. Cancer Lett. 2016;370(1):85–90. https://doi.org/10.1016/j.canlet.2015.10.009 PMID: 26477683
17. Branca RT, Cleveland ZI, Fubara B, Kumar CSSR, Maronpot RR, Leuschner C, et al. Molecular MRI for sensitive and specific detection of lung metastases. Proc Natl Acad Sci U S A. 2010;107(8):3693–7. https://doi.org/10.1073/pnas.1000386107 PMID: 20142483
18. Andrini LB, Marin GH, Inda AM, Bruzzoni Giovanelli H, Garcia M, Errecalde JO, et al. Anti-tumoral effect of a cell penetrating and interfering peptide targeting PP2A/SET interaction. 2020.
19. Barh D. Omics Approaches in Breast Cancer: Towards Next-Generation Diagnosis, Prognosis and Therapy. Springer; 2014.
20. Clarke R. Animal models of breast cancer: their diversity and role in biomedical research. Breast Cancer Res Treat. 1996;39(1):1–6. https://doi.org/10.1007/BF01806073 PMID: 8738601
21. Walrath JC, Hawes JJ, Van Dyke T, Reilly KM. Genetically engineered mouse models in cancer research. Adv Cancer Res. 2010;106:113–64. https://doi.org/10.1016/S0065-230X(10)06004-5 PMID: 20399958
22. Martin DS, Balis ME, Fisher B, Frei E, Freireich EJ, Heppner GH, et al. Role of murine tumor models in cancer treatment research. Cancer Res. 1986;46(4 Pt 2):2189–92. PMID: 3948188
23. Kohnken R, Porcu P, Mishra A. Overview of the use of murine models in leukemia and lymphoma research. Front Oncol. 2017;7:22. https://doi.org/10.3389/fonc.2017.00022 PMID: 28265553
24. Sanmamed MF, Chester C, Melero I, Kohrt H. Defining the optimal murine models to investigate immune checkpoint blockers and their combination with other immunotherapies. Ann Oncol. 2016;27(7):1190–8. https://doi.org/10.1093/annonc/mdw041 PMID: 26912558
25. Guerin MV, Finisguerra V, Van den Eynde BJ, Bercovici N, Trautmann A. Preclinical murine tumor models: a structural and functional perspective. Elife. 2020;9:e50740. https://doi.org/10.7554/eLife.50740 PMID: 31990272
26. Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. Nat Rev Cancer. 2007;7(9):659–72. https://doi.org/10.1038/nrc2193 PMID: 17721431
27. Hurst J, Maniar N, Tombarkiewicz J, Lucas F, Roberson C, Steplewski Z, et al. A novel model of a metastatic human breast tumour xenograft line. Br J Cancer. 1993;68(2):274–6. https://doi.org/10.1038/bjc.1993.327 PMID: 8394103
28. Fantozzi A, Christofori G. Mouse models of breast cancer metastasis. Breast Cancer Res. 2006;8(4):212. https://doi.org/10.1186/bcr1530 PMID: 16887003
29. Khanna C, Hunter K. Modeling metastasis in vivo. Carcinogenesis. 2005;26(3):513–23.
30. Okano M, Oshi M, Butash A, Okano I, Saito K, Kawaguchi T, et al. Orthotopic implantation achieves better engraftment and faster growth than subcutaneous implantation in breast cancer patient-derived xenografts. J Mammary Gland Biol Neoplasia. 2020;25(1):27–36. https://doi.org/10.1007/s10911-020-09442-7 PMID: 32109311
31. Yoshida GJ. Applications of patient-derived tumor xenograft models and tumor organoids. J Hematol Oncol. 2020;13(1):4. https://doi.org/10.1186/s13045-019-0829-z PMID: 31910904

32. Dobrolecki LE, Airhart SD, Alferez DG, Aparicio S, Behbod F, Bentires-Alj M, et al. Patient-derived xenograft (PDX) models in basic and translational breast cancer research. Cancer Metastasis Rev. 2016;35(4):547–73. https://doi.org/10.1007/s10555-016-9653-x PMID: 28025748
33. Weigelt B, Peterse JL, van 't Veer LJ. Breast cancer metastasis: markers and models. Nat Rev Cancer. 2005;5(8):591–602. https://doi.org/10.1038/nrc1670 PMID: 16056258
34. Fisher B, Bauer M, Wickerham DL, Redmond CK, Fisher ER, Cruz AB, et al. Relation of number of positive axillary nodes to the prognosis of patients with primary breast cancer. An NSABP update. Cancer. 1983;52(9):1551–7. https://doi.org/10.1002/1097-0142(19831101)52:9<1551::aid-cncr2820520902>3.0.co;2-3 PMID: 6352003
35. Kim JB, O'Hare MJ, Stein R. Models of breast cancer: is merging human and animal models the future? Breast Cancer Res. 2004;6(1):22–30. https://doi.org/10.1186/bcr645 PMID: 14680482
36. El-Abd E, Sultan A, Shalaby E, Matalkah F. Animal Models of Breast Cancer. Omics Approaches in Breast Cancer. Springer; 2014. p. 297–314.
37. 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. J Cereb Blood Flow Metab. 2020;40(9):1769–77. https://doi.org/10.1177/0271678X20943823 PMID: 32663096
38. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. J Pharmacol Pharmacother. 2010;1(2):94–9. https://doi.org/10.4103/0976-500X.72351 PMID: 21350617
39. Team D. DATAtab: online statistics calculator. DATAtab eU Graz, Austria. 2023. Available from: https://datatabnet.
40. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measure. 1960;20(1):37–46. https://doi.org/10.1177/001316446002000104
41. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–74. PMID: 843571
42. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods. 2021;12(1):55–61. https://doi.org/10.1002/jrsm.1411 PMID: 32336025
43. Nakagawa S, Lagisz M, O'Dea RE, Pottier P, Rutkowska J, Senior AM, et al. orchaRd 2.0: an R package for visualizing meta-analyses with orchard plots. 2023.
44. Duval S, Tweedie R. A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. J Am Stat Assoc. 2000;95(449):89. https://doi.org/10.2307/2669529
45. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455–63. https://doi.org/10.1111/j.0006-341x.2000.00455.x PMID: 10877304
46. Wiencke JK. Impact of race/ethnicity on molecular pathways in human cancer. Nat Rev Cancer. 2004;4(1):79–84. https://doi.org/10.1038/nrc1257 PMID: 14708028
47. Rhee J, Han S-W, Oh D-Y, Kim JH, Im S-A, Han W, et al. The clinicopathologic characteristics and prognostic significance of triple-negativity in node-negative breast cancer. BMC Cancer. 2008;8:307. https://doi.org/10.1186/1471-2407-8-307 PMID: 18947390
48. Ihemelandu CU, Leffall LD Jr, Dewitty RL, Naab TJ, Mezghebe HM, Makambi KH, et al. Molecular breast cancer subtypes in premenopausal and postmenopausal African-American women: age-specific prevalence and survival. J Surg Res. 2007;143(1):109–18. https://doi.org/10.1016/j.jss.2007.03.085 PMID: 17950079
49. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer. 2007;109(9):1721–8. https://doi.org/10.1002/cncr.22618 PMID: 17387718
50. Haffty BG, Yang Q, Reiss M, Kearney T, Higgins SA, Weidhaas J, et al. Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. J Clin Oncol. 2006;24(36):5652–7. https://doi.org/10.1200/JCO.2006.06.5664 PMID: 17116942
51. Curado MP. Breast cancer in the world: incidence and mortality. Salud Publica Mex. 2011;53(5):372–84. PMID: 22218791
52. King RJB, Robins MW. Cancer biology. Pearson Education; 2006.
53. Ban KA, Godellas CV. Epidemiology of breast cancer. Surg Oncol Clin N Am. 2014;23(3):409–22. https://doi.org/10.1016/j.soc.2014.03.011 PMID: 24882341
54. de las Mulas JM, Reymundo C. Animal models of human breast carcinoma: canine and feline neoplasms. Rev Oncología. 2000;2(6):274–81. https://doi.org/10.1007/bf02979590
55. Cheng L, Ramesh AV, Flesken-Nikitin A, Choi J, Nikitin AY. Mouse models for cancer stem cell research. Toxicologic Pathol. 2010;38(1):62–71.
56. Fridriksdottir AJR, Petersen OW, Rønnov-Jessen L. Mammary gland stem cells: current status and future challenges. Int J Dev Biol. 2011;55(7–9):719–29. https://doi.org/10.1387/ijdb.113373af PMID: 22161829
57. Steele VE, Lubet RA, Moon RC. Preclinical animal models for the development of cancer chemoprevention drugs. Cancer chemoprevention. Springer; 2005. p. 39-46.
58. Sharpless NE, Depinho RA. The mighty mouse: genetically engineered mouse models in cancer drug development. Nat Rev Drug Discov. 2006;5(9):741–54. https://doi.org/10.1038/nrd2110 PMID: 16915232
59. Olson B, Li Y, Lin Y, Liu ET, Patnaik A. Mouse models for cancer immunotherapy research. Cancer Discov. 2018;8(11):1358–65. https://doi.org/10.1158/2159-8290.CD-18-0044 PMID: 30309862

60. Coffelt SB, de Visser KE. Immune-mediated mechanisms influencing the efficacy of anticancer therapies. Trends Immunol. 2015;36(4):198–216. https://doi.org/10.1016/j.it.2015.02.006 PMID: 25857662
61. Talmadge JE, Singh RK, Fidler IJ, Raz A. Murine models to evaluate novel and conventional therapeutic strategies for cancer. Am J Pathol. 2007;170(3):793–804. https://doi.org/10.2353/ajpath.2007.060929 PMID: 17322365
62. Soliman N, Rice ASC, Vollert J. A practical guide to preclinical systematic review and meta-analysis. Pain. 2020;161(9):1949–54. https://doi.org/10.1097/j.pain.0000000000001974 PMID: 33449500
63. Tricco AC, Brehaut J, Chen MH, Moher D. Following 411 Cochrane protocols to completion: a retrospective cohort study. PLoS One. 2008;3(11):e3684. https://doi.org/10.1371/journal.pone.0003684 PMID: 18997866
64. 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
65. Ferreira GS, Veening-Griffioen DH, Boon WPC, Moors EHM, Gispen-de Wied CC, Schellekens H, et al. Correction: a standardised framework to identify optimal animal models for efficacy assessment in drug development. PLoS One. 2019;14(7):e0220325. https://doi.org/10.1371/journal.pone.0220325 PMID: 31329650
66. van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. Can animal models of disease reliably inform human studies? PLoS Med. 2010;7(3):e1000245. https://doi.org/10.1371/journal.pmed.1000245 PMID: 20361020
67. Godlee F. We need better animal research, better reported. British Medical Journal Publishing Group; 2018.
68. Peers IS, South MC, Ceuppens PR, Bright JD, Pilling E. Can you trust your animal study data? Nat Rev Drug Discov. 2014;13(7):560. https://doi.org/10.1038/nrd4090-c1 PMID: 24903777
69. Pound P, Ritskes-Hoitinga M. Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. J Transl Med. 2018;16(1):304. https://doi.org/10.1186/s12967-018-1678-1 PMID: 30404629
70. Shultz LD, Ishikawa F, Greiner DL. Humanized mice in translational biomedical research. Nat Rev Immunol. 2007;7(2):118–30. https://doi.org/10.1038/nri2017 PMID: 17259968
71. Fidler IJ, Kripke ML. The challenge of targeting metastasis. Cancer Metastasis Rev. 2015;34(4):635–41. https://doi.org/10.1007/s10555-015-9586-9 PMID: 26328524
72. Dranoff G. Experimental mouse tumour models: what can be learnt about human cancer immunology? Nat Rev Immunol. 2011;12(1):61–6. https://doi.org/10.1038/nri3129 PMID: 22134155
73. Kalinowski L, Saunus JM, McCart Reed AE, Lakhani SR. Breast cancer heterogeneity in primary and metastatic disease. Breast Cancer Metastasis and Drug Resistance. 2019. p. 75–104.
74. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to meta-analysis: John Wiley & Sons; 2021.
75. Stanley TD. Beyond Publication Bias. J Econ Surv. 2005;19(3):309–45. https://doi.org/10.1111/j.0950-0804.2005.00250.x
76. Hönekopp J, Linden AH. Heterogeneity estimates in a biased world. PLoS One. 2022;17(2):e0262809. https://doi.org/10.1371/journal.pone.0262809 PMID: 35113897
77. Jackson D. The implications of publication bias for meta-analysis' other parameter. Stat Med. 2006;25(17):2911–21. https://doi.org/10.1002/sim.2293 PMID: 16345059
78. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. Stat Med. 2007;26(25):4544–62. https://doi.org/10.1002/sim.2889 PMID: 17476644
79. Ioannidis JPA, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ. 2007;176(8):1091–6. https://doi.org/10.1503/cmaj.060410 PMID: 17420491