RESEARCH ARTICLE

The Efficacy of Trastuzumab in Animal Models of Breast Cancer: A Systematic Review and Meta-Analysis

Jiarong Chen[1,2], Canhong Yang[3], Bin Guo[4], Emily S. Sena[5], Malcolm R. Macleod[5], Yawei Yuan[2]*, Theodore C. Hirst[5]*

1 Department of Oncology, Jiangmen Central Hospital, Affiliated Jiangmen Hospital of Sun Yat-Sen University, Jiangmen, Guangdong 529030, P. R. China, 2 Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, 1838 Guangzhou DaDao Bei, Guangzhou, Guangdong 510515, P. R. China, 3 Department of Neurology, the Third Affiliated Hospital of Southern Medical University, Guangzhou 510630, P. R. China, 4 Department of Hepatobiliary Surgery, Zhujiang Hospital, Southern Medical University, Guangzhou 510282, P. R. China, 5 Centre for Clinical Brain Sciences, Chancellors Building, University of Edinburgh, 49 Little France Crescent, Edinburgh EH16 4SB, United Kingdom

* theodore.hirst@ed.ac.uk

Abstract

Background

Breast cancer is the most frequent cancers and is the second leading cause of cancer death among women. Trastuzumab is an effective treatment, the first monoclonal antibody directed against the human epidermal growth factor receptor 2 (HER2). To inform the development of other effective treatments we report summary estimates of efficacy of trastuzumab on survival and tumour volume in animal models of breast cancer.

Methods

We searched PubMed and EMBASE systematically to identify publications testing trastuzumab in animal models of breast cancer. Data describing tumour volume, median survival and animal features were extracted and we assessed quality using a 12-item checklist. We analysed the impact of study design and quality and evidence for publication bias.

Results

We included data from 83 studies reporting 169 experiments using 2076 mice. Trastuzumab treatment caused a substantial reduction in tumour growth, with tumours in treated animals growing to 32.6% of the volume of tumours in control animals (95%CI 27.8%-38.2%). Median survival was prolonged by a factor of 1.45 (1.30–1.62). Many study design and quality features accounted for between-study heterogeneity and we found evidence suggesting publication bias.

Conclusion

We have found trastuzumab to be effective in animal breast cancer models across a range of experimental circumstances. However the presence of publication bias and a low prevalence of measures to reduce bias provide a focus for future improvements in preclinical breast cancer research.

Introduction

Breast cancer is the second most common cancer in the world. With an estimated 1.67 million new cancer cases diagnosed in 2012, breast cancer is the most frequent cancer and second leading cause of cancer death among women [1]. Treatment options including surgery, radiation therapy, and systemic therapy have improved the survival and quality of life of breast cancer patients over the past 30 years [2]. Targeted therapies, which selectively focus on specific molecular targets on tumour cells, provide potential to improve survival and quality of life and to reduce drug side effects [3, 4].

The human epidermal growth factor receptor-2 (HER2), overexpressed in between 15% and 20% of breast cancers [5], is one of the most common molecular targets for targeted therapies [2]. Described as a proto-oncogene, HER2 expression in human cancer was first observed during screening of DNA samples derived from breast cancers [6]. HER2 gene amplification significantly predicts both reduced overall survival and shorter time to relapse in patients with breast cancer. Higher expression of HER2 is associated with a more aggressive disease course leading to worse clinical outcomes [6] as well as adverse prognostic features including advanced pathologic stage [7] and number of metastatic axillary lymph nodes [8]. As a predictor of risk of occurrence and response to therapy HER2 testing is recommended in pathology examination for all newly diagnosed invasive breast cancer and first recurrences of breast cancer [5, 9] as part of to inform staging and choice of therapy.

Trastuzumab, also known as Herceptin, is the first monoclonal antibody directed against HER2. Following studies in tumour xenografts showing that murine monoclonal antibody (MAb) 4D5 inhibits the proliferation of human breast cancer cells overexpressing the HER2 receptor [10], trastuzumab was created in 1990 by humanizing the 4D5 mouse antibody to overcome immunogenicity issues [11]. Subsequent first-generation clinical trials demonstrated prolonged survival and time to progression in patients with HER2-positive metastatic breast cancer [12]. As such, trastuzumab has been widely promoted and recommended for HER2-positive breast cancer in combination with chemotherapies such as cisplatin, doxorubicin or paclitaxel [13–15]. Several studies reported improved disease-free and overall survival following 1 year of trastuzumab administered concomitantly or sequentially as an adjuvant to chemotherapy [16]. Furthermore, trastuzumab shows some efficacy when administrated as a single agent in first-line treatment [17].

Trastuzumab has been used clinically for breast cancer since the mid-1990s and experience in humans has been summarised in systematic reviews and meta-analyses [18, 19]. However, efficacy in animal models has not been well characterised and the extent to which animal studies reliably informed early clinical trials is unclear. Animal experiments improve our understanding of disease mechanisms and they are invaluable for clinical research, ultimately informing decisions on when to proceed to clinical trial. Systematic review and meta-analysis of such studies may aid in the selection of the most promising treatment strategies for future clinical trials, as well as in the identification of potential sources of bias arising from limitations

in study design [20]. In the context of treatments known to be effective in humans, a deeper understanding of these strengths and weaknesses of *in vivo* studies may inform the development of new therapies. The Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies (CAMARADES) group have shown—for a range of experimental neurological diseases including glioma, stroke and multiple sclerosis—that reporting of measures to reduce risk of bias (such as randomisation and blinding) is poor and that publication bias is frequently present. These factors lead to a significant overstatement of perceived treatment efficacy [21–23].

Therefore, we aimed to undertake a systematic review of trastuzumab in animal models of breast cancer, to summarise the effect of trastuzumab on survival and tumour volume. We also sought to describe the impact on observed efficacy of study quality features such as randomization and blinding and experimental design features such as drug dose, timing of treatment, species, study quality, so that we might ascertain the value of such studies in predicting clinical efficacy and suggest how future animal research might be streamlined to predict clinical efficacy more reliably.

Materials and Methods

The study was conducted in accordance with a protocol, first published online in December 2014 at http://www.dcn.ed.ac.uk/camarades/research.html#protocols.

Search strategy

We searched PubMed and EMBASE on 4 July 2013, using the terms relating to (breast tumor OR breast cancer) AND (trastuzumab OR Herceptin) (S1 Table). On 18 September 2014, we added three further terms to expand the search with previous names of trastuzumab: (breast tumor OR breast cancer) AND ("rhuMAb HER2" OR "Anti-p185HER2 Monoclonal Antibody" OR "muMAb 4D5"). For EMBASE the search was limited to animals, and for PubMed we used a validated animal search filter [24]. Studies in all languages were accepted.

Study selection

Two investigators independently screened all titles and abstracts to identify studies meeting the inclusion criteria. The full text of all these potentially eligible studies were retrieved and independently assessed for eligibility by two investigators (Canhong Yang and Bin Guo), disagreements were discussed with a third investigator (Jiarong Chen) after full text reading. We included controlled studies that reported the efficacy of trastuzumab monotherapy in laboratory animals with induced breast cancer. We included studies reporting outcome as either median survival or tumour volume in treatment and control groups. Publications had to state the number of animals per group, and we excluded Reviews, Books, Letters, Clinical trials, Case reports, or Editorials.

Data extraction

We extracted data to the CAMARADES data manager (Microsoft Access) for assessment of study quality and evidence synthesis. Extracted information included: article information (title, author, journal, publication year); agent for treatment and control group; trastuzumab treatment (dose, total dose and frequency, route of administration, delay to treatment); tumour volume data for treatment and control group (original tumour size included, measurement method and frequency, measurement day after treatment, measurement day after tumour implantation), and whether the study reported the blinded assessment of outcome. We

recorded data from the last time point for tumour volume measurement for treatment and control group; median survival data for treatment and control group; experimental animals (number of animals for every group, species and strain, age, sex and original weight), breast cancer model (breast cancer cell type, HER2 expression level (as determined by in situ hybridization (ISH), immunology and histology chemistry (IHC), western blot or by other means), trastuzumab resistance, tumour implantation method and implantation site, number or volume of implanted tumour cells, use of estradiol (dose, route of administration, delay to treatment). Missing data were requested from study authors. We extracted trastuzumab dose as mg/kg. Where this was quoted as μg, we estimated a weight-adjusted dose based on a typical average mouse weight of 25g.

Risk of bias

Two review authors independently assessed the risk of bias in included studies by considering the following characteristics: peer-reviewed publication, and reporting of random allocation of tumour-bearing animal to treatment or control groups, blinded assessment of outcome, sample size calculation, statement of potential conflict of interests, compliance with animal welfare policy, explanation of rationale for disease model used, or multiple tumour models used, standardised number or volume of tumour cells implanted, reported number of animals in which the tumour did not grow, reporting and explanation of excluded animals, presentation of evidence that trastuzumab acts directly against the tumour, and consistent implantation site. We collated these into a quality checklist, similar to that described previously [21].

Analysis

1) Volume data. We initially intended to log transform the dataset (to account for exponential tumour growth) and then generate a normalized mean difference summary statistic (see protocol). However, through late 2014-early 2015 (after the data extraction phase) we observed that this produced nonparametrically distributed data and, with a large number of studies reporting improvements greater than +90%, there was a resultant skew with a large number of studies clustered close to the ceiling efficacy of +100%. This effect persisted even if the data were log transformed as planned *a priori*. Following discussion we instead decided to use a volume ratio summary statistic, as when log-transformed this would produce parametrically distributed data without ceiling. We elected to adopt the method suggested by Higgins et al [25]. We have expressed volume ratio values as the inverse log of these summary estimates to make the data easier to understand, and present these such that values over 1 favour the treatment group to make interpretation of figures more intuitive.

We used DerSimonian and Laird random effects meta-analysis to pool studies and stratified meta-analysis to search for sources of heterogeneity. For each group we again present the inverse log to give more intuitive values for mean and SE. As we have stratified the volume data 13 times, we used Bonferroni correction to give a critical value for significance of $p < 0.00385$.

2) Survival data. We generated an effect score of median survival ratio by dividing the treated group by control group. We then log-transformed the data and used a modified form of DerSimonian and Laird meta-analysis (weighted by the number of animals in the study instead of inverse variance) to generate a global efficacy estimate [26]. We used stratified meta-analysis as above.

3) Publication bias. We used funnel plots, Egger regression and Trim and Fill analysis [27] to search for evidence of publication bias.

Results

Identification of papers

We identified 1651 publications (PubMed 741, EMBASE 910), with 1299 articles remaining after duplicate removal. After title and abstract screening, full texts of 269 articles were retrieved. We found 85 studies reporting outcome as either median survival or tumour volume, but one study was excluded because tumour size was only inferred by measuring the activity of secreted molecules in the blood; tumour volume was measured directly in other studies. A further experiment was excluded during the analysis stage as the control group was treated with androstenedione, an androgen known to affect breast tumour growth in its own right [28]. Thus we included 83 studies reporting 169 experiments using 2076 mice in our meta-analysis (Fig 1). No experiments used other species (See S2 Table).

We have generated a spreadsheet containing the raw data extracted from each publication and published this on Figshare for public access (https://figshare.com/articles/Trastuzumab_preclinical_SRMA_-_raw_dataset/3406513).

Of these, 158 experiments assessed tumour volume as the outcome measure, with 11 reporting animal survival. Trastuzumab treatment was associated with a substantial reduction in tumour growth—with control tumours growing to 3.07 times the volume of treated tumours (95%CI 2.62–3.60). That is, tumour volume in treated animals was reduced by 67.4% (95%CI 61.8% - 72.2%). We observed considerable between-study heterogeneity (χ^2 = 11376, df = 157, p<0.00385; I^2 = 98.6%). Median survival was prolonged by a factor of 1.45 (1.30–1.62) although we did not observe any between-study heterogeneity in this smaller dataset (χ^2 = 6.0, df = 10, p>0.05). Consequently we did not conduct a stratified meta-analysis of survival data.

Risk of bias

Study quality was evaluated based on a modified 12-point checklist [21] with a median quality score of 6 (IQR 5–7; range 3 to 9. S3 Table). All of the 83 papers were published in a peer-reviewed journal, 36 (43%) reported random allocation to group, 36 (43%) stated whether a potential conflict of interest existed, and 31 (37%) described compliance with animal welfare policy. 49 (59%) studies compared more than one model, and 69 (83%) presented evidence that trastuzumab acts directly against the tumour. Standardized number or volume of tumour cells implanted was reported in 78 (94%) studies, and implantation site was consistent in 71 (86%) of studies. The number of animals in which the tumour did not grow was reported in only 4 (5%) studies and explanation of excluded animals was recorded in 12 (14%) studies. Surprisingly, none of the studies described the blinded outcome assessment or reported a sample size calculation. Stratification by the number of study quality checklist items scored accounted for significant heterogeneity between studies (χ^2 = 1373, df = 6, p<0.00385), with higher quality studies reporting lower efficacy (Fig 2a). Similarly, studies that randomized group allocation reported lower efficacy (χ^2 = 68.1, df = 1, P<0.00385. Fig 2b).

We also found substantial evidence of publication bias indicative of an overestimation of observed efficacy and implying an absence of small, inefficacious studies in the literature. Inspection of a funnel plot for tumour volume data suggests an excess of imprecise studies reporting high efficacy; there were only 14 experiments with results favouring control and of these, 9 reported volume ratios between 0.9 and 1 (Fig 2c). Egger regression was again consitent with publication bias (t = 8.772, p<0.001; Fig 2d) and Trim and Fill analysis suggests the presence of 38 'missing studies' (red plots, Fig 2c). After inclusion of these imputed missing studies the adjusted global efficacy estimate was reduced from 3.07 to 1.77 (95%CI 1.46–2.14). For survival studies, the funnel plot did not show any obvious asymmetry (Fig 2e), although the
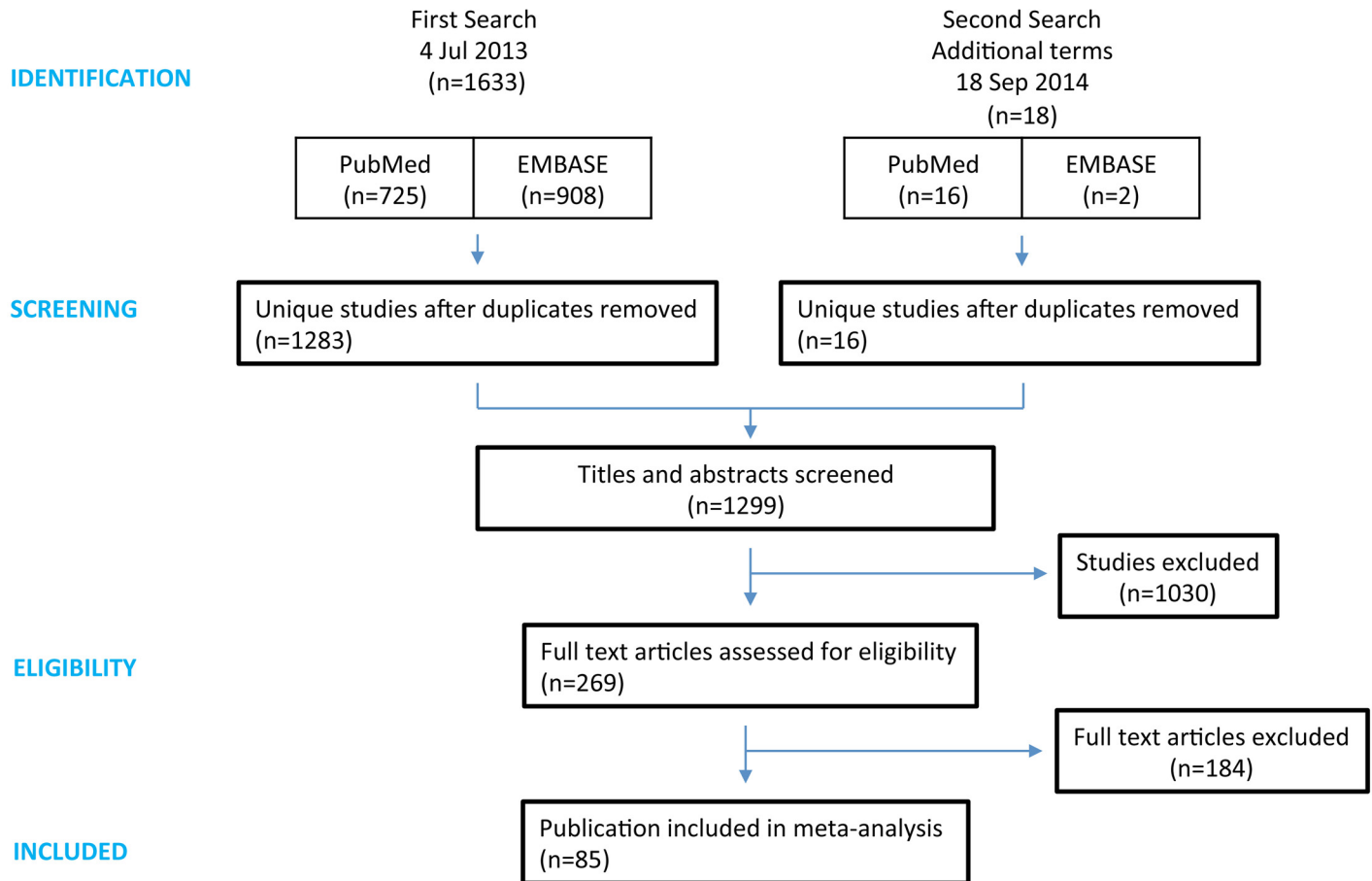
Fig 1. PRISMA flowchart of study selection. The first search was performed on 4 July 2013, using the terms relating to (breast tumor OR breast cancer) AND (trastuzumab OR Herceptin) (See S1 Table). On 18 September 2014, we added three further terms to expand the search with previous names of trastuzumab: (breast tumor OR breast cancer) AND ("rhuMAb HER2" OR "Anti-p185HER2 Monoclonal Antibody" OR "muMAb 4D5").

doi:10.1371/journal.pone.0158240.g001

presence of publication bias was suggested by a positive intercept on Egger regression (t = 5.340, p<0.001; Fig 2f). Trim and fill did not impute any missing studies for survival data. Importantly, we would expect some of the studies reported here to be neutral. For instance, some studies tested effciacy in cells known to have limited or no sensitivity to trastuzumab. The presence of such studies is therefore likely to lead to an underestimation of the scale of publication bias.

We have performed a cumulative meta-analysis for tumour volume to assess whether earlier studies might have been more influential in decisions relating to clinical trial design. Plotting global efficacy estimate for all studies published up to and including any given year shows a steady trend for values to become more conservative with time, with for instance a global efficacy estimate of 10.4 (4.01–27.1) in 2000 (Fig 3). This suggests a possible overestimation of efficacy during the earlier years of trastuzumab research, which may be due to differences in study design or quality measures. An alternative explanation would be if early studies tested efficacy in situations where we would expect it to be higher, for instance using tumour cell lines more sensitive to trastuzumab, or with intratumoural injection of drug. However, when we limit the analysis to tumour cell lines graded as ++ or +++ we see the same non-significant trend of decline in efficacy (S1 Fig)
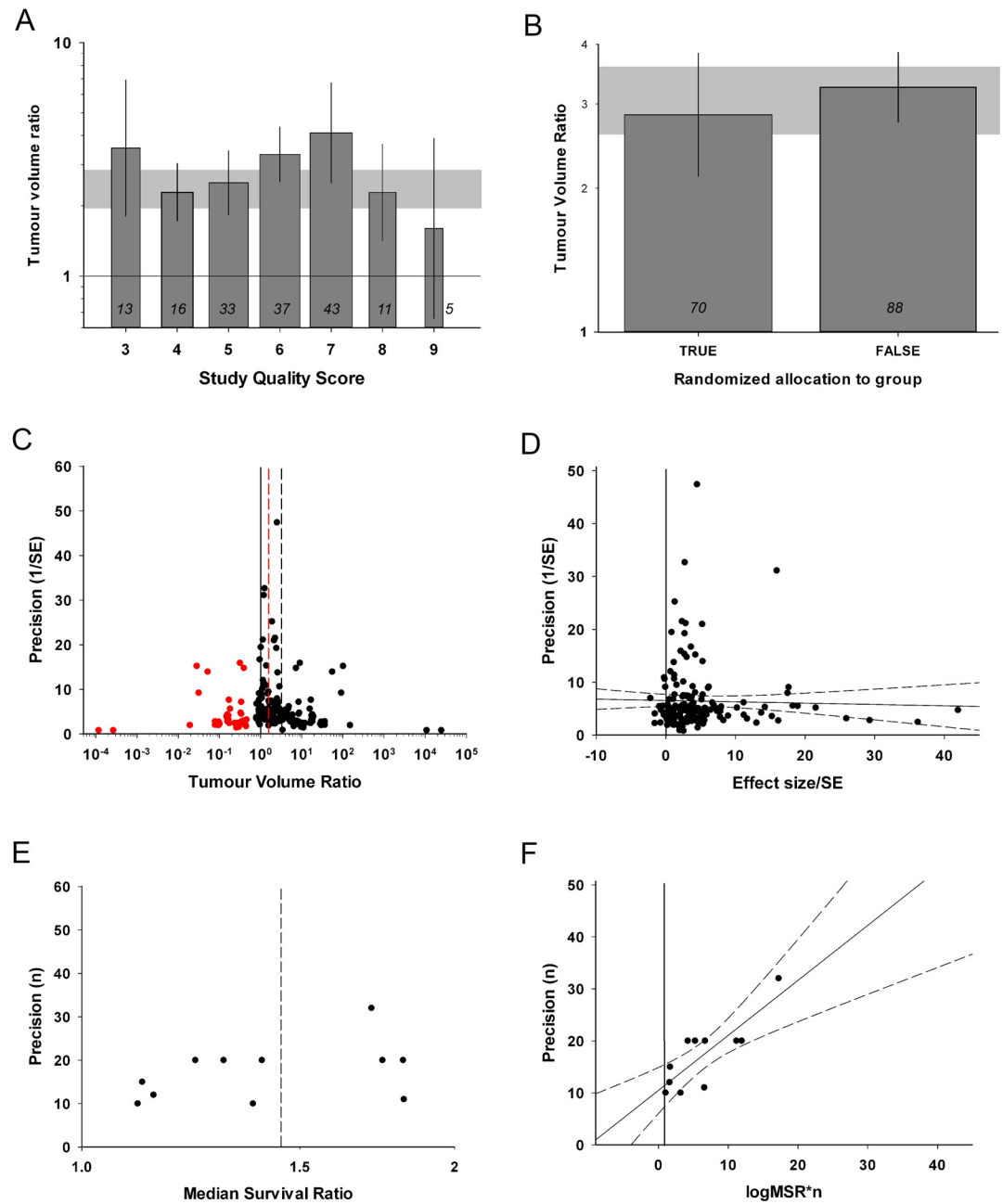
Fig 2. Evaluation of study quality and publication bias. (A) Quality score is associated with between-study heterogeneity tumour volume studies, with a trend for higher quality studies to report smaller effect sizes ($\chi^2 = 1373$, df = 6, p<0.00385). (B) Randomized studies were associated with smaller tumour volume reductions than non-randomised studies ($\chi^2 = 68.1$, df = 1, p<0.00385). (C and D) Both funnel plot and Egger regression indicate a presence of publication bias in tumour volume data (t = 8.772, p<0.001). Trim and Fill analysis added 38 'missing' studies, with a sizeable reduction in global efficacy estimate (red plots). (E and F) For survival studies, funnel plot does not show any obvious asymmetry, but Egger regression suggests a publication bias (t = 8.772, p<0.001). The grey bands in A and B represent global 95% confidence intervals; columns represent mean ± 95%CI and column width a measure of number of comparisons within each stratum; the number of comparisons is included at the base of each column. Dotted lines in C and E represent global estimates of efficacy before (grey) and after (red) Trim and fill analysis. Dotted lines in D and F represent 95%CI of the regression. The vertical solid lines in A, C, D and F represent the level of neutral treatment effect.

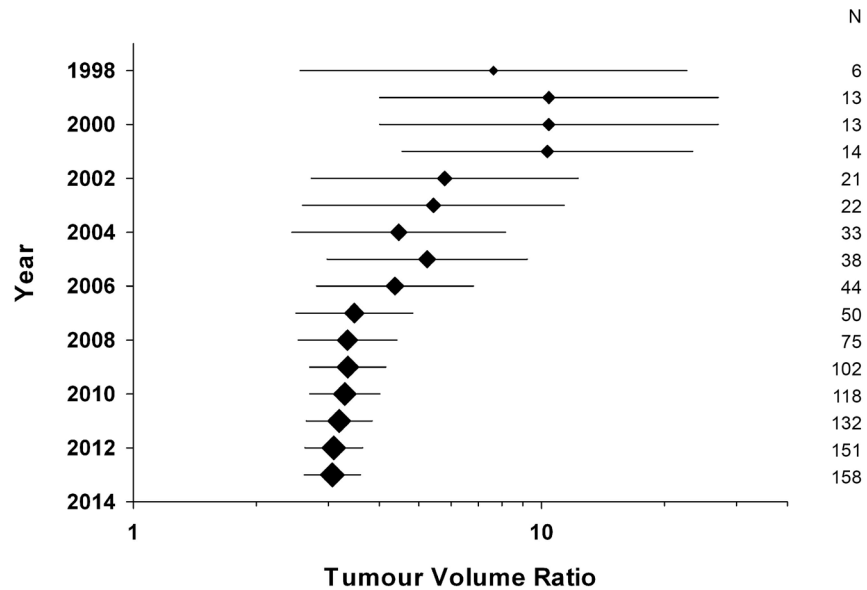doi:10.1371/journal.pone.0158240.g002

Fig 3. Cumulative meta-analysis. Plotting cumulative global efficacy estimate with time shows a steady trend for values to become more conservative with time. Plot size is representative of the number of studies, values for which are included on the right hand side, and error bars represent the 95%CI of the mean.

doi:10.1371/journal.pone.0158240.g003

Animal and tumour models

The experimental breast cancer tumour models fell into three broad categories: tumour xeno-graft inoculation (146/158 volume studies, 11/11 survival studies), tumours induced using mouse mammary tumour virus (MMTV, 10/158), oncomice with a genetic predisposition to develop spontaneous tumours (1/158); and in 1 experiment the type of tumour was unclear. Efficacy was higher in experiments using oncomice and tumours induced by MMTV than those using xenografts ($\chi^2 = 291$, df = 3, p<0.00385). We observed 24 different tumour xenograft models, of which the BT-474 cell line was the most commonly used in both tumor volume (53/146) and survival studies (5/11). The specific xenograft model used were associated with a significant proportion of the observed heterogeneity in this group ($\chi^2 = 4416$, df = 23, p<0.00385. Fig 4a) and the MDA-MB-453 model appeared to be most sensitive to trastuzumab.

Tumours with augmented HER2 expression were commonly used (142/158 volume experiments, 11/11 survival). Overexpression of HER2 was linked to a better outcome ($\chi^2 = 160$, df = 1, p<0.00385, Fig 4b) when compared to HER2 negative or low expression. Furthermore, we found that studies reporting the use of tumour cells known to be trastuzumab-resistant *in vitro* were associated with smaller effect sizes ($\chi^2 = 1009$, df = 1, p<0.00385). Tumours were inoculated either subcutaneously (88/158 volume studies, 8/11 survival) or into mammary tissue (66/158; 3/11), and there was a greater reduction in tumour volume with subcutaneous inoculation than for tumours induced by intramammary inoculation ($\chi^2 = 810$, df = 2, p<0.00385; Fig 4c).

All studies included in our analysis used mice to evaluate trastuzumab efficacy. Almost all used mice with altered immune status (155/158 volume studies, 11/11 survival studies), most frequently athymic mice (for both volume (93/158) and survival studies (7/11)), with the next most common using severe combined immunodeficient (SCID) models (23/158 volume, 3/11 survival). These models differed in the observed efficacy ($\chi^2 = 5130$, df = 9, P<0.00385, Fig 4d),
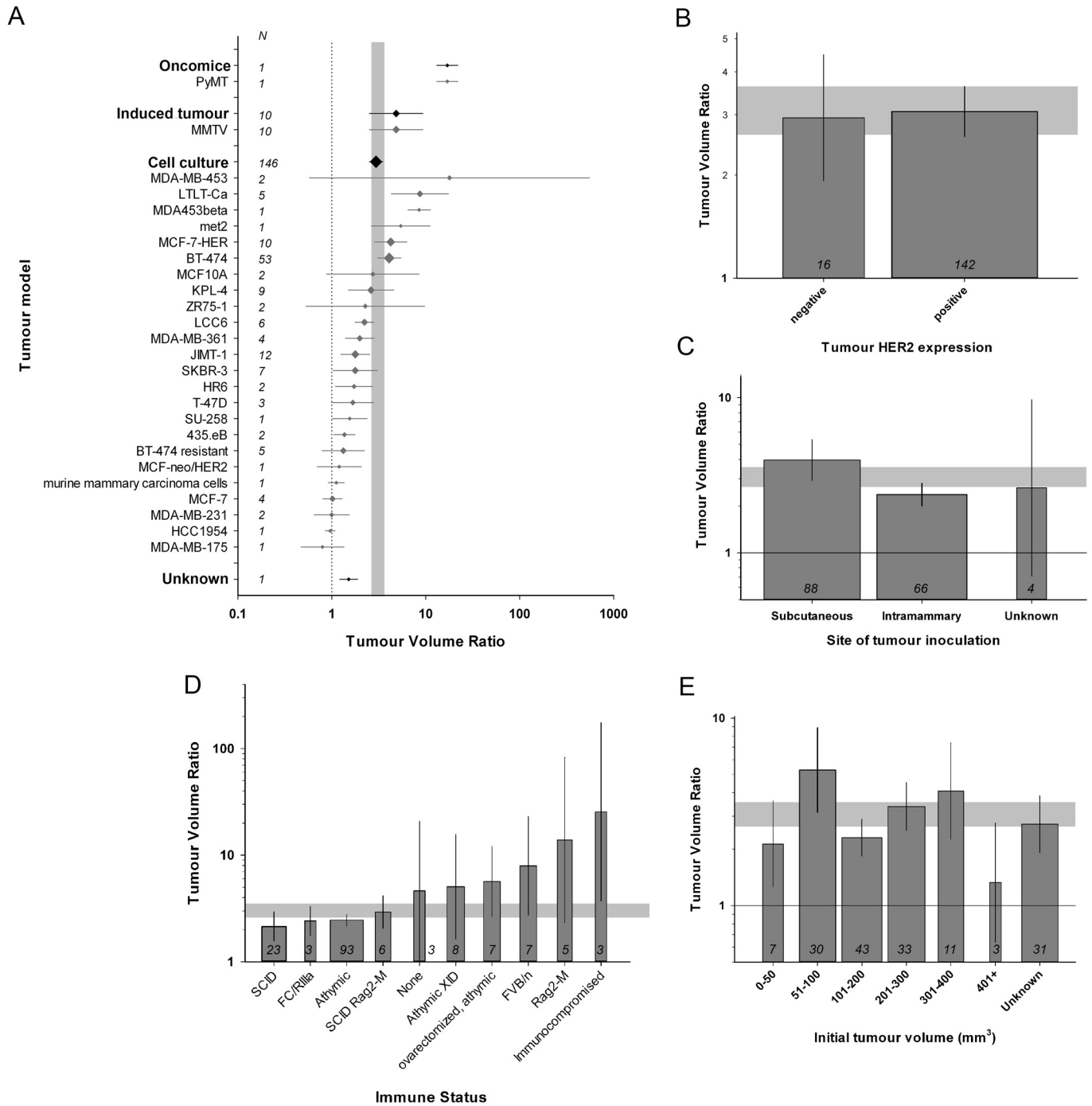
Fig 4. Stratification by animal and tumour model characteristics. (A) The specific xenograft model used were associated with a significant proportion of the between-study heterogeneity (χ^2 = 4416, df = 23, P<0.00385). (B) Overexpression of HER2 in breast cancer tumours was associated with greater efficacy (χ^2 = 160, df = 1, P<0.00385). (C) A greater reduction in tumour volume was observed with subcutaneous inoculation compared to intramammary route (χ^2 = 810, df = 2, P<0.00385). (D) There is comorbidity related between-study heterogeneity (χ^2 = 5130, df = 9, P<0.00385). The more commonly used athymic and SCID mice were associated with lower efficacies. (E) A significant proportion of between-study heterogeneity is accounted for by the tumour volume at treatment initiation, with a smaller treatment effect seen at either extreme (χ^2 = 1042, df = 6, P<0.00385). The grey bands represent global 95% confidence intervals; diamonds (A) and columns (B-E) represent mean±95% CI and diamond size (A) column width (B-E) a measure of number of comparisons within each stratum; the number of comparisons is included inside the y-axis (A) and at the base of each column (B-E). The solid lines in A, C and E represent the level of neutral treatment effect.

doi:10.1371/journal.pone.0158240.g004

with athymic and SCID mice being associated with lower efficacies than animals with non-specific 'immunocompromised' mice or Rag2-M phenotype and those with no reported comorbidity. We included ovarectomy within this comorbidity stratification and efficacy was higher in athymic mice that were ovarectomized than in athymic mice that were not (see Fig 4d).

We used tumour volume at initiation of therapy as a surrogate for the delay to trastuzumab treatment, stratifying studies into groups of 0–50, 51–100, 101–200, 201–300, 301–400 and >400 mm^3, as well as unknown. For studies providing this information, the majority started treatment when tumour volume was between 51 and 300 mm^3. A significant proportion of between-study heterogeneity in volume data is accounted for by this stratification, with a smaller treatment effect seen at either extreme (χ^2 = 1042, df = 6, p<0.00385, Fig 4e). This information was only available for 6 survival studies.

Treatment regimen

To study the influence of different trastuzumab dosing regimens we stratified studies by total trastuzumab dose, by route of delivery and by tumour volume at treatment initiation as discussed above. We did not include treatment duration, administration frequency, number of cycles or other treatment details because specifying optimal dosing regimens in animals was not an aim of this study.

We stratified the total trastuzumab dose into groups of <11, 11–25, 26–50, 51–75, 76–100 and >100 mg/kg, as well as unknown. Higher doses of trastuzumab were associated with greater efficacy (χ^2 = 723, df = 6, P<0.00385, Fig 5a). We observed heterogeneity between different routes of trastuzumab delivery; intraperitoneal injection was most frequently used (130/ 158 volume and 9/11 survival studies). While all were effective, the single study reporting subcutaneous injection reported greater reductions in tumour volume than pooled estimates for other routes (χ^2 = 62, df = 3, p<0.00385, Fig 5b).

Around half of the studies concomitantly implanted intradermal sustained-release estradiol pellets before xenograft inoculation (80/158 tumour volume studies, 4/11 survival studies), and these found greater reduction of tumour volume (χ^2 = 336, df = 1, p<0.00385, Fig 5c). We also found that the choice of control was associated with between-study heterogeneity, in that studies reporting the use of non-functioning antibodies as control were associated with more conservative estimates of efficacy than those using vehicle or non-treated controls (χ^2 = 956, df = 3, P<0.00385, Fig 5d).

Discussion

In this review of 83 publications involving 169 experiments and 2076 animals, trastuzumab monotherapy appears to reduce tumour volume and prolong survival in animals bearing breast tumours. We have identified a large degree of between-study heterogeneity within the volume dataset but not across the smaller number of survival studies; and we have identified a number of features pertaining to both study quality and design that are associated with between-study heterogeneity. We have identified the presence of publication bias in both datasets.

In the tumour volume dataset we have demonstrated substantial between-study heterogeneity; furthermore, all the variables we tested accounted for between-heterogeneity in stratified meta-analysis. Our strategy is limited in its univariate nature, and does not account for covariance between the variables. [22, 29]. Colinearity might be taken into account using multivariate techniques such as a multi-variable meta-regression but this approach is seldom used in meta-analyses of animal data and its validity in this context is not known. The stratification accounting for the largest degree of heterogeneity (and therefore the first candidate variable for a multi-variable approach) was the tumour model used; the large number of models (27, of which 16
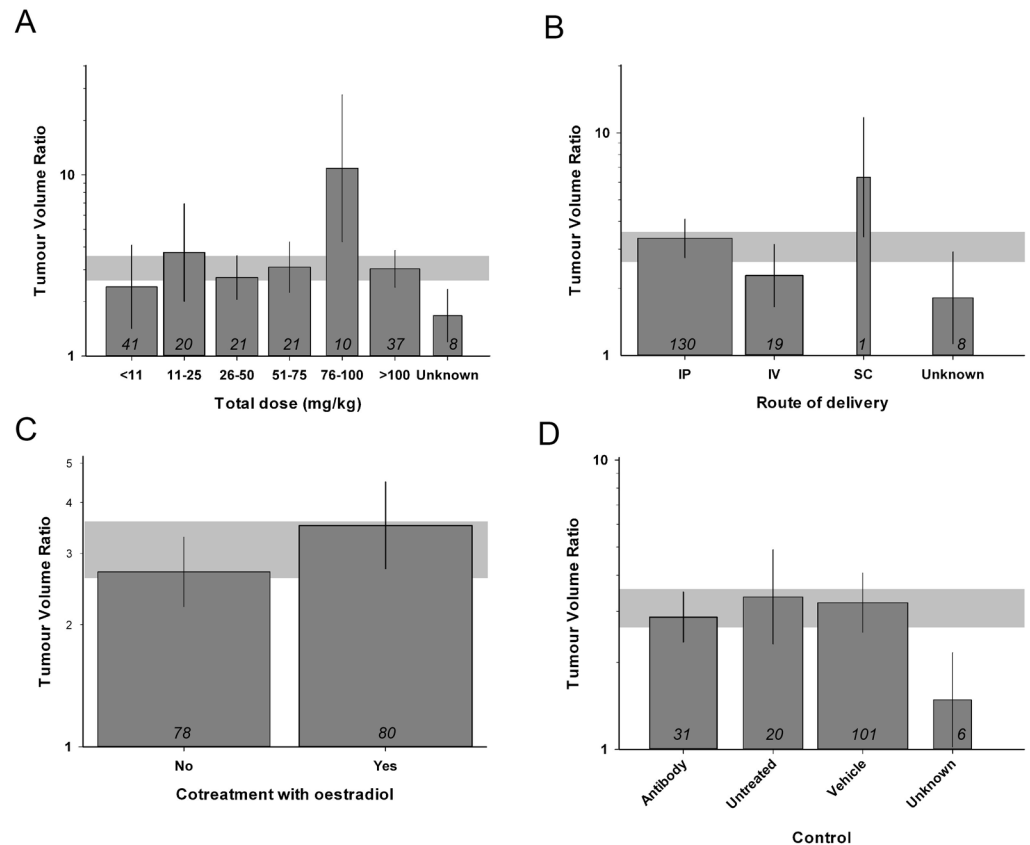
Fig 5. Stratification by trastuzumab treatment characteristics. Total trastuzumab dose (A. $\chi^2 = 723$, df = 6, P<0.00385), route of drug delivery (B. $\chi^2 = 62$, df = 3, P<0.00385), cotreatment with oestradiol (C. $\chi^2 = 336$, df = 1, P<0.00385) and the choice of control (D. $\chi^2 = 956$, df = 3, P<0.00385), accounted for significant between-study variability in reduction of tumour volume. The grey band represents global 95% confidence intervals; columns represent mean ± 95% CI and column width a measure of number of comparisons within each stratum; the number of comparisons is included at the base of each column.

doi:10.1371/journal.pone.0158240.g005

were only used in 3 experiments or fewer) precludes multivariate analysis. We did not see between-study heterogeneity in the analysis of survival experiments. This may be due to the small number of studies, each of small size with low power, or may relate to limitations in the analysis brought about by the absence of data denoting within-study variance [26]. Consequently the results from our analysis should be considered hypothesis generating only and must be interpreted with caution.

We believe the preclinical trastuzumab literature is at high risk of bias. Study quality is modest, with a median score of 6 out of a possible 12 and no study met more than 9 checklist items. Less than half the studies reported randomized group allocation and no studies reported the blinded assessment of outcome or the use of a sample size calculation. We have observed lower efficacy in those studies at lower risk of bias—namely studies achieving a higher quality score and those randomizing—and this phenomenon is well documented in the preclinical neurosciences literature [22, 23]. Furthermore there is considerable publication bias amongst both datasets with a grossly asymmetrical funnel plot for tumour volume data and positive Egger's test for both datasets. Even Trim and Fill, a conservative technique [30], led to an adjusted estimate of efficacy almost 50% lower than the unadjusted figure, implying a large excess of small studies reporting large efficacy in the published literature.

Optimising study quality is important as overstated results from low quality animal studies can misinform decisions to proceed to clinical trial [31] and inaccurate conclusions from these data will inevitably lead to unnecessary animal suffering. There is increasing awareness of this phenomenon and quality is in general improving with time [32]. Nonetheless we believe that a standardized report pattern including measures to reduce risk of bias must be widely adopted for animal research to continue to be valid as a translational strategy [33]. An exemplar reporting framework is the ARRIVE guidelines [33, 34].

The first article evaluating the efficacy of trastuzumab on breast cancer animal model was published in 1991 [35]. However, this article was not included in our analysis as the primary outcome was tumour weight and was therefore did not meet our inclusion criteria. The first positive clinical study of trastuzumab showed objective responses in 43 of 46 patients with tumours overexpressing HER2 in 1996 [36]; but the earliest animal study included in our analysis was published in 1998. There is a trend for reported effect size to become more conservative with time, implying that the problem of overstated efficacy was larger at the very time when this information was most likely to be applied to human trials. Efficacy estimates appear to decrease in magnitude from 2001: this may coincide with attempts to characterise trastuzumab resistance, a process utilising different animal models, dosing regimens and trastuzumab-resistant xenografts [37].

There are several factors that may influence an experimenter's choice of tumour model [38, 39]. We have found that this choice was the largest source of heterogeneity and, when taken with similar findings in glioma studies [21, 23], makes model choice a major determinant of study outcome in cancer studies. This was why we included a justification for tumour model choice or the testing of two or more tumour models as a study quality criterion. BT-474 and MCF-7 (HER2-transfected) were the most popular cell lines and these appeared to be amongst the more trastuzumab-sensitive.

We have also shown that the use of immunocompromised mice is the default *modus operandi* for preclinical breast cancer studies. While this is clearly important for the successful inoculation of xenografts it should be borne in mind that the immune system plays a key role in the body's response to cancer and contributes substantially to the therapeutic effects of monoclonal antibodies, [40, 41] this will not be recapitulated in studies using immunodeficient animals. Greater volume reduction was seen in tumours inoculated subcutaneously than those into mammary fat pad. Since the host environment in the latter more accurately represents that in human disease we recommend that future studies should adopt this approach where possible.

As for control reagents, we noticed a slightly reduced efficacy in studies controlling with antibody (mainly with non-specific immunoglobulin), compared to vehicle and untreated controls. While it may be possible for non-functioning antibodies to modulate the body's response to tumour this observation may also reflect the association between study quality and outcome alluded to above.

In human disease, amplification of HER2 is usually associated with worse prognosis in breast cancer; and trastuzumab is generally more effective against tumours overexpressing HER2 to a higher degree. It is therefore relevant to assess HER2 status in order to prognosticate treatment [42, 43]. This is generally done by measuring HER2 gene copy number with in situ hybridization (ISH) or immunohistochemistry (IHC) [5, 9]. In our analysis we dichotomised HER2 expression status into negative, with expression reported as low or negative, and positive, with HER2 overexpressed, observing greater efficacy in tumours overexpressing HER2. We had initially intended to compare efficacy between different degrees of HER2 overexpression, however we found that studies used a range of different methods to quantify this,

including but not limited to IHC and ISH, and in 65 experiments the degree of HER2 overex-pression was not stated. Therefore this stratification was not possible.

The timing and dosage of treatment for any condition are inherently important and a wide range of nuanced regimens were reported in the literature. However, characterizing details of this were not the primary aims of this study. We have included analyses of total dose and delay to treatment to gain a superficial insight into the regimens used and to give some simple comparisons between types. We have observed what seems to be a dose-response relationship with trastuzumab although side effect profile has not been accounted for due to the lack of information reported in the studies. Tumour volume reduction was relatively consistent with moderate tumour initial volumes, with lower efficacy with larger tumours. To go further than this would require prospective experiments directly comparing specific treatment regimens at various stages of disease. Animal studies in general used intraperitoneal drug delivery, differing from the intravenous route favoured in clinical practice. We saw greater efficacy in the single study reporting subcutaneous delivery; subcutaneous administration has been tested in humans with the advantage of convenience [44, 45], although there may be more serious adverse effects than with intravenous delivery [46].

Hormone level plays an essential role in initiation, promotion, and progression of breast cancer [47, 48]. Estrogen is required for the normal development of the mammary gland and is predominantly produced in the ovary. Several studies reported the use of ovarectomy to simulate the postmenopausal, oestrogen-deplete, state. In these studies all animals were also athymic and trastuzumab efficacy was greater in this group than in the athymic (non-ovarectomized) animals (see Fig 4d and S2 Table). There is some evidence from animal models showing that estrogens are breast carcinogens [49, 50] and when inoculating tumour xeongrafts, concordant oestrogen delivery could increase the chance of tumour growth [51]. Estrogen was used as a cotreatment in about half of the comparisons in our analyses while establishing breast cancer model, mostly via a subcutaneous injection of sustained release estradiol less than a week prior to tumour cell inoculation. Cotreatment resulted a higher efficacy in tumour volume studies. This may reflect variations in tumour model or the fact that tumours co-expressing HER2 and estrogen receptors are associated with a better outcome than those expressing HER2 alone [52]

To summarise, we have found trastuzumab to be effective in animal breast cancer models across a range of experimental circumstances, correlating with clinical experience. There is evidence to suggest this literature is at high risk of bias, through the presence of publication bias and low prevalence of randomization and blinding. We have also defined a number of study design features that appear to affect the perceived efficacy of trastuzumab. This meta-analysis is of course limited, especially in the context of low-quality underlying data and the likely presence of colinearity, so these results should be interpreted with caution. We hope that our conclusions may help to streamline future research in experimental breast cancer and in the wider preclinical setting.

Supporting Information

S1 Fig. Cumulative meta-analysis limited to tumour lines with high HER2 expression.
(TIF)

S1 Table. Search strategy summary.
(DOCX)

S2 Table. Study characteristics.
(DOCX)

S3 Table. Study quality scores.
(DOCX)

S4 Table. PRISMA checklist.
(DOC)

S5 Table. Raw data.
(XLS)

S1 Text. Study protocol.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: JC MM YY TH. Performed the experiments: JC CY BG TH. Analyzed the data: TH. Contributed reagents/materials/analysis tools: JC ES MM TH. Wrote the paper: JC ES MM YY TH.

References

1. DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. CA: a cancer journal for clinicians. 2011; 61(6):409–18. doi: 10.3322/caac.20134 PMID: 21969133.

2. Kawalec P, Lopuch S, Mikrut A. Effectiveness of Targeted Therapy in Patients With Previously Untreated Metastatic Breast Cancer: A Systematic Review and Meta-Analysis. Clinical breast cancer. 2014. doi: 10.1016/j.clbc.2014.10.006 PMID: 25441421.

3. Fojo T, Parkinson DR. Biologically targeted cancer therapy and marginal benefits: are we making too much of too little or are we achieving too little by giving too much? Clinical cancer research: an official journal of the American Association for Cancer Research. 2010; 16(24):5972–80. doi: 10.1158/1078-0432.CCR-10-1277 PMID: 21169250.

4. van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature. 2008; 452(7187):564–70. doi: 10.1038/nature06915 PMID: 18385730.

5. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2013; 31(31):3997–4013. doi: 10.1200/JCO.2013.50.9984 PMID: 24101045.

6. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science. 1987; 235(4785):177–82. PMID: 3798106.

7. Seshadri R, Firgaira FA, Horsfall DJ, McCaul K, Setlur V, Kitchen P. Clinical significance of HER-2/neu oncogene amplification in primary breast cancer. The South Australian Breast Cancer Study Group. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 1993; 11 (10):1936–42. PMID: 8105035.

8. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science. 1989; 244(4905):707–12. PMID: 2470152.

9. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Archives of pathology & laboratory medicine. 2014; 138(2):241–56. doi: 10.5858/arpa.2013-0953-SA PMID: 24099077; PubMed Central PMCID: PMCPMC4086638.

10. Kumar R, Shepard HM, Mendelsohn J. Regulation of phosphorylation of the c-erbB-2/HER2 gene product by a monoclonal antibody and serum growth factor(s) in human mammary carcinoma cells. Molecular and cellular biology. 1991; 11(2):979–86. PMID: 1671297; PubMed Central PMCID: PMC359762.

11. Carter P, Presta L, Gorman CM, Ridgway JB, Henner D, Wong WL, et al. Humanization of an anti-p185HER2 antibody for human cancer therapy. Proceedings of the National Academy of Sciences of the United States of America. 1992; 89(10):4285–9. PMID: 1350088; PubMed Central PMCID: PMC49066.

12. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England journal of medicine. 2001; 344(11):783–92. doi: 10.1056/NEJM200103153441101 PMID: 11248153.

13. Pegram MD, Lipton A, Hayes DF, Weber BL, Baselga JM, Tripathy D, et al. Phase II study of receptor-enhanced chemosensitivity using recombinant humanized anti-p185HER2/neu monoclonal antibody plus cisplatin in patients with HER2/neu-overexpressing metastatic breast cancer refractory to chemotherapy treatment. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 1998; 16(8):2659–71. PMID: 9704716.

14. Baselga J, Manikhas A, Cortes J, Llombart A, Roman L, Semiglazov VF, et al. Phase III trial of nonpegylated liposomal doxorubicin in combination with trastuzumab and paclitaxel in HER2-positive metastatic breast cancer. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO. 2014; 25(3):592–8. doi: 10.1093/annonc/mdt543 PMID: 24401928.

15. Tolaney SM, Barry WT, Dang CT, Yardley DA, Moy B, Marcom PK, et al. Adjuvant paclitaxel and trastuzumab for node-negative, HER2-positive breast cancer. The New England journal of medicine. 2015; 372(2):134–41. doi: 10.1056/NEJMoa1406281 PMID: 25564897; PubMed Central PMCID: PMC4313867.

16. Perez EA, Romond EH, Suman VJ, Jeong JH, Davidson NE, Geyer CE Jr, et al. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2011; 29(25):3366–73. doi: 10.1200/JCO.2011.35.0868 PMID: 21768458; PubMed Central PMCID: PMC3164242.

17. Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, et al. Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2002; 20 (3):719–26. PMID: 11821453.

18. Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, et al. Trastuzumab containing regimens for early breast cancer. The Cochrane database of systematic reviews. 2012; 4:CD006243. doi: 10.1002/14651858.CD006243.pub2 PMID: 22513938.

19. Jiang Z, Yang Y, Pan Z, Yue Z, Li L. [Efficacy of lapatinib versus trastuzumab in neoadjuvant therapy of HER-2 positive breast cancer: a meta-analysis]. Zhonghua yi xue za zhi. 2014; 94(48):3854–8. PMID: 25623322.

20. Sena ES, Currie GL, McCann SK, Macleod MR, Howells DW. Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism. 2014; 34(5):737–42. doi: 10.1038/jcbfm.2014.28 PMID: 24549183; PubMed Central PMCID: PMC4013765.

21. Hirst TC, Vesterinen HM, Sena ES, Egan KJ, Macleod MR, Whittle IR. Systematic review and meta-analysis of temozolomide in animal models of glioma: was clinical efficacy predicted? Br J Cancer. 2013; 108(1):64–71. doi: 10.1038/bjc.2012.504 PMID: 23321511; PubMed Central PMCID: PMCPMC3553514.

22. Sena ES, Briscoe CL, Howells DW, Donnan GA, Sandercock PA, Macleod MR. Factors affecting the apparent efficacy and safety of tissue plasminogen activator in thrombotic occlusion models of stroke: systematic review and meta-analysis. Journal of cerebral blood flow and metabolism: official journal of the International Society of Cerebral Blood Flow and Metabolism. 2010; 30(12):1905–13. doi: 10.1038/jcbfm.2010.116 PMID: 20648038; PubMed Central PMCID: PMC3002882.

23. Hirst TC, Vesterinen HM, Conlin S, Egan KJ, Antonic A, Lawson McLean A, et al. A systematic review and meta-analysis of gene therapy in animal models of cerebral glioma: why did promise not translate to human therapy? 2015. doi: 10.1002/ebm2.6

24. Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Laboratory animals. 2010; 44(3):170–5. doi: 10.1258/la.2010.009117 PMID: 20551243; PubMed Central PMCID: PMC3104815.

25. Higgins JP, White IR, Anzures-Cabrera J. Meta-analysis of skewed data: combining results reported on log-transformed or raw scales. Statistics in medicine. 2008; 27(29):6072–92. doi: 10.1002/sim.3427 PMID: 18800342; PubMed Central PMCID: PMC2978323.

26. Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, et al. Meta-analysis of data from animal studies: a practical guide. J Neurosci Methods. 2014; 221:92–102. doi: 10.1016/j.jneumeth. 2013.09.010 PMID: 24099992.

27. 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. PMID: 10877304.

28. Dimitrakakis C, Bondy C. Androgens and the breast. Breast Cancer Res. 2009; 11(5):212. doi: 10. 1186/bcr2413 PMID: 19889198; PubMed Central PMCID: PMCPMC2790857.

29. Watzlawick R, Sena ES, Dirnagl U, Brommer B, Kopp MA, Macleod MR, et al. Effect and reporting bias of RhoA/ROCK-blockade intervention on locomotor recovery after spinal cord injury: a systematic review and meta-analysis. JAMA Neurol. 2014; 71(1):91–9. doi: 10.1001/jamaneurol.2013.4684 PMID: 24297045.

30. Simonsohn U, Nelson LD, Simmons JP. p-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results. Perspect Psychol Sci. 2014; 9(6):666–81. doi: 10.1177/1745691614553988 PMID: 26186117.

31. de Vries RB, Wever KE, Avey MT, Stephens ML, Sena ES, Leenaars M. The usefulness of systematic reviews of animal experiments for the design of preclinical and clinical studies. ILAR J. 2014; 55 (3):427–37. doi: 10.1093/ilar/ilu043 PMID: 25541545; PubMed Central PMCID: PMCPMC4276599.

32. 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. eCollection 2015 Oct. PMID: 26460723

33. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. A call for transparent reporting to optimize the predictive value of preclinical research. Nature. 2012; 490(7419):187–91. doi: 10.1038/nature11556 PMID: 23060188; PubMed Central PMCID: PMCPMC3511845.

34. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS biology. 2010; 8(6):e1000412. doi: 10. 1371/journal.pbio.1000412 PMID: 20613859; PubMed Central PMCID: PMCPMC2893951.

35. Shepard HM, Lewis GD, Sarup JC, Fendly BM, Maneval D, Mordenti J, et al. Monoclonal antibody therapy of human cancer: taking the HER2 protooncogene to the clinic. J Clin Immunol. 1991; 11(3):117–27. PMID: 1679763.

36. Baselga J, Tripathy D, Mendelsohn J, Baughman S, Benz CC, Dantis L, et al. Phase II study of weekly intravenous recombinant humanized anti-p185HER2 monoclonal antibody in patients with HER2/neu-overexpressing metastatic breast cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 1996; 14(3):737–44. PMID: 8622019.

37. Albanell J, Baselga J. Unraveling resistance to trastuzumab (Herceptin): insulin-like growth factor-I receptor, a new suspect. J Natl Cancer Inst. 2001; 93(24):1830–2. PMID: 11752000.

38. Borowsky AD. Choosing a mouse model: experimental biology in context—the utility and limitations of mouse models of breast cancer. Cold Spring Harb Perspect Biol. 2011; 3(9):a009670. doi: 10.1101/ cshperspect.a009670 PMID: 21646376; PubMed Central PMCID: PMCPMC3181037.

39. Pegram M, Ngo D. Application and potential limitations of animal models utilized in the development of trastuzumab (Herceptin): a case study. Adv Drug Deliv Rev. 2006; 58(5–6):723–34. doi: 10.1016/j. addr.2006.05.003 PMID: 16876287.

40. Adam JK, Odhav B, Bhoola KD. Immune responses in cancer. Pharmacol Ther. 2003; 99(1):113–32. PMID: 12804702.

41. Bianchini G, Gianni L. The immune system and response to HER2-targeted treatment in breast cancer. The Lancet Oncology. 2014; 15(2):e58–68. doi: 10.1016/S1470-2045(13)70477-7 PMID: 24480556.

42. Cheng H, Bai Y, Sikov W, Sinclair N, Bossuyt V, Abu-Khalaf MM, et al. Quantitative measurements of HER2 and phospho-HER2 expression: correlation with pathologic response to neoadjuvant chemotherapy and trastuzumab. BMC Cancer. 2014; 14:326. doi: 10.1186/1471-2407-14-326 PMID: 24885187; PubMed Central PMCID: PMCPMC4037428.

43. Nuciforo P, Thyparambil S, Aura C, Garrido-Castro A, Vilaro M, Peg V, et al. High HER2 protein levels correlate with increased survival in breast cancer patients treated with anti-HER2 therapy. Mol Oncol. 2016; 10(1):138–47. doi: 10.1016/j.molonc.2015.09.002 PMID: 26422389.

44. Fallowfield L, Osborne S, Langridge C, Monson K, Kilkerr J, Jenkins V. Implications of subcutaneous or intravenous delivery of trastuzumab; further insight from patient interviews in the PrefHer study. Breast. 2015; 24(2):166–70. doi: 10.1016/j.breast.2015.01.002 PMID: 25623753.

45. Ismael G, Hegg R, Muehlbauer S, Heinzmann D, Lum B, Kim SB, et al. Subcutaneous versus intravenous administration of (neo)adjuvant trastuzumab in patients with HER2-positive, clinical stage I-III breast cancer (HannaH study): a phase 3, open-label, multicentre, randomised trial. The Lancet Oncology. 2012; 13(9):869–78. doi: 10.1016/S1470-2045(12)70329-7 PMID: 22884505.

46. Anon. Subcutaneous trastuzumab (Herceptin). Ready to use, but more serious adverse effects than intravenous (IV) trastuzumab. Prescrire Int. 2015; 24(159):93. PMID: 25941697.

47. Stanislawska-Sachadyn A, Sachadyn P, Limon J. Transcriptomic Effects of Estrogen Starvation and Induction in the MCF7 Cells. The Meta-analysis of Microarray Results. Curr Pharm Biotechnol. 2015. PMID: 26511976.

48. Reznikov A. Hormonal impact on tumor growth and progression. Exp Oncol. 2015; 37(3):162–72. PMID: 26422099.

49. Bocchinfuso WP, Korach KS. Mammary gland development and tumorigenesis in estrogen receptor knockout mice. J Mammary Gland Biol Neoplasia. 1997; 2(4):323–34. PMID: 10935020.

50. Castagnetta L, Granata OM, Cocciadiferro L, Saetta A, Polito L, Bronte G, et al. Sex steroids, carcinogenesis, and cancer progression. Ann N Y Acad Sci. 2004; 1028:233–46. doi: 10.1196/annals.1321.028 PMID: 15650249.

51. Pickar JH, Komm BS. Selective estrogen receptor modulators and the combination therapy conjugated estrogens/bazedoxifene: A review of effects on the breast. Post Reprod Health. 2015; 21(3):112–21. doi: 10.1177/2053369115599090 PMID: 26289836.

52. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin Med Res. 2009; 7(1–2):4–13. doi: 10.3121/cmr.2009.825 PMID: 19574486; PubMed Central PMCID: PMCPMC2705275.