Can contemporary trials of chemotherapy for HER2-negative metastatic breast cancer detect overall survival benefit?

Sherko Kümmel1
Christian Jackisch2
Volkmar Müller3
Andreas Schneweiss4
Sandra Klawitter5
Michael P Lux6

1Breast Unit Essen, Kliniken Essen-Mitte, Essen, Germany; 2Department of Obstetrics and Gynecology, Sana Klinikum Offenbach, Offenbach, Germany; 3Department of Gynecology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 4Division of Gynecologic Oncology, National Center for Tumor Diseases, University Hospital, Heidelberg, Germany; 5Medical Affairs – Biostatistics and Epidemiology, Roche Pharma AG, Grenzach-Wyhlen, Germany; 6Department of Gynecology and Obstetrics, Erlangen University Hospital, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany

Purpose: Although several trials have demonstrated improved progression-free survival (PFS) with first-line regimens for HER2-negative metastatic breast cancer (mBC), overall survival (OS) benefit is elusive. We calculated required sample sizes to power for OS using published data from recent mBC trials.

Patients and methods: Randomized superiority trials of first-line chemotherapy/targeted therapy for HER2-negative mBC including >150 patients, meeting the primary efficacy objective, and published in 2000–2018 were identified. The sample sizes required to power for PFS and OS were calculated retrospectively for each trial using observed results and study/recruitment follow-up durations (α=0.05, two-sided log-rank test, 80% power), and summarized as a factor (x) relative to actual sample size.

Results: Nine of 13 identified trials reported all information required for retrospective sample size calculation. Six had sample sizes larger than required to demonstrate a significant PFS benefit but all would have required larger sample sizes to demonstrate significant OS benefit with the observed results. In ten trials, the required sample size was ≥5-fold larger to power for OS than PFS.

Conclusion: Designing trials to test potential new treatments for HER2-negative mBC is challenging, requiring a balance of regulatory acceptability, feasibility, and realistic medical assumptions to calculate sample sizes. Powering for OS is particularly difficult in heterogeneous populations with long postprogression survival, potential crossover, heterogeneous poststudy therapy, and evolving treatment standards. Validated surrogate endpoints are critical. Ongoing trials of cancer immunotherapy (new mode of action) in triple-negative mBC (more homogeneous, shorter OS and postprogression survival, fewer treatment options) may show a new pattern.

Keywords: progression-free survival, overall survival, endpoint, metastatic breast cancer, clinical trial, regulatory authorities

Introduction

In metastatic breast cancer (mBC), selection of the most appropriate endpoint for clinical trials is becoming increasingly important when evaluating new first-line therapies. In HER2-positive mBC, for which a number of targeted agents exist, several trials across treatment settings have demonstrated overall survival (OS) benefits from HER2-directed therapies.1 In HER2-negative mBC, however, where the target is less clear and patient selection is more challenging, progression-free survival (PFS) benefits have rarely translated into statistically significant OS benefits. To date, no Phase III trial evaluating antiangiogenic agents, cyclin-dependent kinase 4/6 inhibitors, mTOR inhibitors, or poly(adenosine diphosphate–ribose) polymerase inhibitors has shown a
statistically significant OS improvement. OS is considered an unambiguous endpoint and is the global gold standard for demonstrating clinical benefit. However, extending life is not necessarily valuable if accompanied by significant quality of life (QoL) deterioration. Other disadvantages of OS as a primary endpoint are bias caused by treatment evolution during long studies, the diluting effect of crossover, numerous heterogeneous subsequent treatment lines, and the need for large patient numbers and/or long follow-up before obtaining results. This is particularly problematic in first-line trials, in which patients typically receive multiple treatment lines after progression. Consequently, authorities including the European Medicines Agency accept PFS as a relevant endpoint and approve drugs based on PFS benefit.

The correlation between PFS and OS appears to be less robust in settings with longer postprogression survival and/or effective subsequent therapies, whereas in later treatment lines, the likelihood of showing an OS benefit increases. However, a recent analysis of 40 randomized controlled trials in HER2-negative hormone receptor-positive mBC indicated a significant association between PFS/time to progression (TTP) and OS, irrespective of treatment line. To explore this topic further, we used published data from contemporary HER2-negative mBC trials to calculate the sample sizes required to power for OS compared with sample sizes actually used. Based on our findings, we discuss the challenges of designing trials in HER2-negative mBC, where powering for OS is sometimes unrealistic, unfeasible, or unfundable, with the aim of improving future trial planning and design.

Design
Clinical trials were identified from a systematic search of MEDLINE (details in Table S1) using the following criteria: randomized superiority trials; first-line chemotherapy or targeted therapy for HER2-negative mBC; >150 patients; meeting the primary efficacy objective (“positive” trials); and published in English in a peer-reviewed journal between January 1, 2000 and February 15, 2018.

The sample sizes required to power for PFS/TTP and OS were calculated retrospectively for each trial using the observed median PFS/TTP and median OS in the treatment groups for treatment effect, the actual recruitment period, and the actual total study duration (α=0.05, two-sided log-rank test, 80% power). Dropout rates were not considered for sample size calculation. nQuery Advisor (version 7.0; Statistical Solutions Ltd, Cork, Ireland) was used for sample size calculations. If information on the total study duration was missing, we chose a simple pragmatic assumption that the study period was one-third longer than the recruitment duration.

The retrospectively calculated sample sizes were summarized as a factor (x) relative to the actual sample size. x <1 would require x-fold fewer cases to show a significant benefit, whereas x >1 required x-fold more cases.

Results
Analysis data set
Thirteen trials met the selection criteria (Table 1). Of these, nine reported all information required for retrospective sample size calculation. In four reports (all published before 2006), insufficiently described study duration made it difficult or impossible to understand fully the statistical assumptions for sample size calculation. Only one trial had OS as the primary endpoint.

In most trials, the HRs showed a stronger treatment effect on PFS/TTP than OS (Figure 1). Four trials showed statistically significantly improved OS.

Retrospective sample size calculation
Table 2 shows the retrospectively calculated sample sizes required to show PFS/TTP and OS benefit with the observed data compared with the actual sample sizes. According to these calculations, six of 13 trials had sample sizes larger than required to demonstrate a significant PFS benefit. However, all would have required a larger sample size to demonstrate a significant OS benefit with the observed results. The increase in sample size ranged from 1.2-fold to 2,460-fold. In nine of the 12 trials with OS information, the calculated required sample size to demonstrate a significant OS benefit with the observed OS results was at least fivefold greater than the actual sample size. Figure 2 summarizes the sample size increase required to show a significant OS benefit with the reported data.

In all but one trial, a larger sample size would be required to show OS than PFS benefit. In 10 of the 12 trials with available OS results, the sample size required to power for OS was at least fivefold larger than that needed to power for PFS.

Discussion
Our analyses suggest that in the first-line HER2-negative mBC setting, it is a high hurdle to conduct a trial with adequate power to detect an OS improvement. Sample sizes to power for OS are usually extremely large and substantially larger than required to power for PFS.

The generally larger PFS than OS treatment effect in HER2-negative mBC is consistent with a recently reported
study across various tumor types.\textsuperscript{27} Our findings are also consistent with reports in the literature suggesting that demonstrating an OS benefit is becoming increasingly unrealistic in contemporary clinical trials.\textsuperscript{2} A trial without crossover may answer the question of OS most cleanly. However, if the investigational agent has shown clear activity, the possibility of crossover has to be discussed. An Independent Data Monitoring Committee may feel obliged to stop a trial because of a clear signal, but it will then be impossible to conclude on the secondary endpoint of OS. Furthermore, a second trial of the same agent cannot be conducted after proven benefit because it is difficult to consent patients to be randomized between an experimental agent and a control arm known to be inferior. At times of rapid innovation, endpoints allowing prompt application of therapy optimization to standard clinical care are required. Therefore, it is important to determine whether progression-based endpoints are suitable for demonstrating utility. Available endpoints include PFS, TTP, and time to treatment failure. These allow earlier provision of study results and can be more sensitive indicators of treatment benefit because they are not affected by further treatment lines or crossover.\textsuperscript{24,29} Another benefit is comparability, as PFS is currently the most commonly used primary endpoint in Phase III trials. However, there is no clear evidence that PFS is a surrogate for OS.

In a recent analysis of PFS and OS in 58 randomized Phase II/III trials evaluating first-line systemic therapy for HER2-negative hormone receptor-positive mBC, several factors besides first-line therapy were reported to influence OS.\textsuperscript{30} These included prior endocrine therapy, prior (neo)adjuvant chemotherapy, types and lines of postprogression therapy, as well as disease characteristics associated with prognosis. Geographic region also influenced OS, presumably because of differences in healthcare patterns, management, and access in different countries.

In our analysis, the trial in which the actual sample size and the retrospectively calculated sample size for OS were most similar was IMELDA, a maintenance trial evaluating the addition of capecitabine to maintenance bevacizumab after bevacizumab/taxane induction therapy as a new treatment approach.\textsuperscript{24} In these patients already demonstrating chemotherapy resistance before progression potentially anticipates development of capecitabine sensitivity to induction therapy, switching to capecitabine before progression potentially anticipates development of resistance. In IMELDA, a significant OS benefit was demonstrated with a relatively small sample size but the retrospectively calculated sample size suggested that a larger sample size was needed. This is explained by differences in the methodology used for sample size calculation compared with the trial analysis method. Importantly, sample size calculation is only an estimation. Conversely, trial outcome is not proof and there is 5% error for demonstrating a significant benefit.

There appeared to be a gradual increase in median OS in the investigational arm over time (Table 2). Such cross-trial comparisons have obvious limitations, particularly when including maintenance vs treatment strategies. Nevertheless, median OS with experimental therapy remained <2 years in all trials evaluating chemotherapy alone, crossing the 2-year threshold only with the introduction of targeted therapy (bevacizumab). This presumably reflects not only treatment effect but also earlier diagnosis, better disease management, and an increase in the number of subsequent therapy options available. Indeed, similar increases in median OS can be seen in the control arm.

Given these challenges, how should we test effectiveness most appropriately in the first-line HER2-negative mBC setting? While Health Technology Assessment bodies worldwide accept PFS as a meaningful endpoint for clinical trials, progression-based outcomes are not recognized in Germany by the Institute for Quality and Efficiency in Health Services and the Joint Federal Committee (G-BA). These organizations focus on QoL, safety, OS, and morbidity, whereas PFS alone is not considered a meaningful endpoint, nor (in contrast with the clinical view) as an aspect of morbidity. The rationale for the G-BA's stance is that superior progression-based outcomes evaluated by imaging are not considered to represent relevant benefits for patients. Patient relevance is accepted only if progression is recorded, for example, through symptoms perceptible to the patient. However, guidelines recommend assessing tumor burden every 8 weeks to allow prompt detection of metastatic progression, discontinuation of ineffective treatment with associated side effects, and prevention of tumor-associated symptoms that could be avoided by a change of treatment or strategy.\textsuperscript{31}

Irrespective of surrogacy for OS, many believe that PFS is an important and relevant outcome for patients, associated with improved overall QoL, physical functioning, and emotional well-being.\textsuperscript{32} Extending PFS was ranked as more important than tumor shrinkage, limiting side effects, or treatment frequency in a questionnaire-based survey. Self-rated QoL was the highest after respondents had been told that their disease was responding to treatment. Therefore, progression-based parameters should generally be accepted as patient-relevant endpoints. Furthermore, changing therapy at progression affects patients' lives. A new therapy may be associated with new side effects and/or a new treatment schedule and mode of administration. The consequences
Table 1 Overview of trials included in the analysis. Trials are ordered according to date enrollment began (earliest first)

| Trial           | Recruitment period | Control arm (A)                                                                 | Experimental arm (B)                                                                 | Missing information for retrospective sample size calculation |
|-----------------|--------------------|---------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------|
| HEPI 013        | Sep 1990–Nov 1992  | Cyclophosphamide, methotrexate, fluorouracil                                   | Cyclophosphamide, epirubicin, fluorouracil                                        | Yes                                          |
| SBG 9403        | Feb 1995–Jan 1999  | Epirubicin                                                                      | Vinorelbine + epirubicin                                                            | Yes                                          |
| Jassem et al 2001 | Nov 1996–Apr 1998 | Fluorouracil, doxorubicin, cyclophosphamide                                    | Doxorubicin + paclitaxel                                                           | Yes                                          |
| von Minckwitz et al 2005 | Nov 1996–Sep 2001 | Cyclophosphamide, methotrexate, fluorouracil                                   | Bendamustine, methotrexate, fluorouracil                                          | Yes                                          |
| Bontenbal et al 2005 | Mar 1997–Apr 2002 | Fluorouracil, doxorubicin, cyclophosphamide                                    | Doxorubicin + docetaxel                                                           | No                                           |
| Albain et al 2008 | Aug 1999–Apr 2002 | Paclitaxel                                                                      | Gemcitabine + paclitaxel                                                          | No                                           |
| E2100           | Dec 2001–May 2004  | Paclitaxel                                                                      | Bevacizumab + paclitaxel                                                          | No                                           |
| Sparano et al 2009 | Sep 2004–Nov 2006 | Docetaxel                                                                      | Pegylated liposomal doxorubicin + docetaxel                                         | No                                           |
| RIBBON-1        | Dec 2005–Aug 2007  | Capecitabine                                                                    | Bevacizumab + capecitabine                                                       | No                                           |
| AVADO           | Mar 2006–Apr 2007  | Docetaxel                                                                       | Bevacizumab + docetaxel                                                           | No                                           |
| ATX (BOOG)      | Jun 2007–Dec 2010  | Paclitaxel + bevacizum maintenance                                              | Paclitaxel + bevacizum + capecitabine induction, then bevacizumab and capcitabine maintenance | No                                           |
| IMELDA          | Jul 2009–Mar 2011  | Docetaxel + bevacizum maintenance                                               | Docetaxel + bevacizum induction, then bevacizumab + bevacizum maintenance          | No                                           |
| MERIDIAN        | Aug 2012–Dec 2013  | Paclitaxel                                                                      | Paclitaxel + bevacizum                                                            | No                                           |

Notes: *Total study duration missing. *Data reported for capecitabine cohort (anthracycline/taxane cohort not included because of heterogeneity of chemotherapy backbone). Data reported for comparison of bevacizumab 15 mg/kg vs placebo (bevacizumab 7.5 mg/kg not included in this analysis as there was no significant improvement in the primary endpoint) but events/patients for reported sample size calculations include all three treatment arms as reported in the statistical design section of the publication.

Abbreviations: NR, not reported; OS, overall survival; PFS, progression-free survival; TTP, time to disease progression.

Figure 1 Summary of PFS/TTP and OS HRs across trials. Vertical bars represent 95% CIs, except for MERIDIAN, which shows the 99% CI reported for this coprimary endpoint.

Abbreviations: OS, overall survival; PFS, progression-free survival; TTP, time to progression.
Table 1 Overview of trials included in the analysis. Trials are ordered according to date enrollment began (earliest first).

| Trial | Total no of patients in trial | Observed median, months (arm A vs arm B) | Retrospectively calculated sample size | Factor (x) |
|-------|-------------------------------|------------------------------------------|----------------------------------------|------------|
|       | PFS                           | TTP                                      | OS                                     | OS sample size/N |
|       | (arm A vs arm B)              |                                          |                                        | OS/(PFS/TTP) sample size |
|       | PFS/TTP                       |                                          |                                        |                        |
|       | OS                            |                                          |                                        |                        |
|       | OS (PFS/TTP)                  |                                          |                                        |                        |
|       | OS/TTP                        |                                          |                                        |                        |
|       | OS                            |                                          |                                        |                        |
|       | OS/TTP                        |                                          |                                        |                        |
|       | OS                            |                                          |                                        |                        |
|       | OS/TTP                        |                                          |                                        |                        |

Notes: Median values are not comparable with the other trials, first because PFS and OS were calculated from the time of randomization to maintenance therapy rather than the start of first-line therapy, and second because only patients with response or stable disease after induction therapy were included in the randomized population. Retrospectively calculated sample size (arm A vs arm B).

Abbreviations: OS, overall survival; PFS, progression-free survival; TTP, time to disease progression.

Table 2 Summary of trial outcomes

| Trial                  | Total no of patients in trial | Observed median, months (arm A vs arm B) | Retrospectively calculated sample size | Factor (x) |
|------------------------|-------------------------------|------------------------------------------|----------------------------------------|------------|
|                        | PFS                           | TTP                                      | OS                                     | OS sample size/N |
|                        | PFS/TTP                       |                                          |                                        | OS/(PFS/TTP) sample size |
|                        | OS                            |                                          |                                        |                        |
|                        | OS/TTP                        |                                          |                                        |                        |
|                        | OS                            |                                          |                                        |                        |
|                        | OS/TTP                        |                                          |                                        |                        |
|                        | OS                            |                                          |                                        |                        |
|                        | OS/TTP                        |                                          |                                        |                        |

Notes: Median values are not comparable with the other trials, first because PFS and OS were calculated from the time of randomization to maintenance therapy rather than the start of first-line therapy, and second because only patients with response or stable disease after induction therapy were included in the randomized population. Retrospectively calculated sample size (arm A vs arm B).

Abbreviations: OS, overall survival; PFS, progression-free survival; TTP, time to disease progression.
of disease progression are depressive reactions, grief, and despair. The possibility of tumor control is the most important reason for patients agreeing to systemic therapy. Fear of disease progression is the most commonly reported psychological burden in patients.

We acknowledge that PFS is not a perfect endpoint, potentially being influenced by assessment intervals, choice of target lesions, and measurement technology. Some of these challenges are overcome by Independent Central Review, which is important for accepting PFS as an endpoint. Regarding the limitations of OS, several elegant biostatistical methods have been developed to account for crossover, such as inverse probability of censoring weighting (IPCW) and the rank-preserving structural failure time (RPSFT) model. However, these approaches are not flawless: IPCW assumes that there are no unknown or unmeasured confounding factors that could influence crossover and OS, whereas RPSFT assumes that the effect of treatment is constant across time and/or treatment lines. No single validated standard for statistical correction of crossover has been established in settings with long postprogression survival.

With the increasing use of maintenance therapies, eg, in ovarian cancer, alternatives to PFS and OS have emerged, including intermediate endpoints such as time to second progression or time to first or second subsequent therapy. These endpoints merit consideration in future trial designs in HER2-negative mBC. For trials evaluating endocrine therapies, time to first chemotherapy can also be a valuable endpoint, with clear patient relevance. Alternative endpoints used in other tumor types include quality-adjusted time without symptoms or toxicity and quality-adjusted PFS. However, it is essential that any endpoint is clearly defined and that the precise definition is used consistently across trials measuring the effect of treatment. Changes in molecular markers may also be of interest as surrogate endpoints.

In an attempt to quantify the medical benefits of new drugs, composite scales including pharmacoeconomic parameters have been introduced, such as the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS) and the American Society of Clinical Oncology Value Framework. A recent survey indicated that many trials demonstrating statistically significant improvements in efficacy did not meet the ESMO-MCBS clinical benefit threshold, particularly trials in the palliative setting.

The challenge of large sample sizes required to show OS improvement in clinical trials has been accompanied by increased interest in real-world data (RWD). In some cases, RWD evaluation has suggested improved OS from a treatment despite the lack of OS benefit in prospective randomized clinical trials. The main advantages of RWD are the very large sample sizes available for analysis and inclusion of broader, more heterogeneous patient populations with common comorbidities than is possible in a clinical trial, reflecting populations presenting in routine oncology practice. However, there are many limitations and even with sophisticated statistical methodology, RWD are exposed to important potential biases. Therefore, RWD can be viewed...
only as complementary to randomized clinical trials, not as an alternative.

A limitation of our analysis is the focus on chemotherapy and antiangiogenic agents. Numerous ongoing trials in the first-line HER2-negative mBC setting are evaluating cancer immunotherapy agents, which have a different mode of action and thus may exhibit different effects on PFS and OS. Furthermore, many of these trials focus on triple-negative mBC, a slightly more homogeneous population with shorter OS expectancy, shorter postprogression survival, and fewer treatment options after progression. All of these factors may affect the ability to demonstrate a significant OS effect, and, therefore, the patterns observed in our analysis may not predict future trials of cancer immunotherapy. Interestingly, several ongoing Phase III trials of immunotherapy in triple-negative mBC evaluate OS as the (co)primary endpoint. Another potential criticism is that the proportion of patients completing treatment is not taken into account. This information is missing in some of the publications, particularly in the older trials, but may have an impact on outcomes.

Conclusion

Although there are many reasons why OS is an attractive endpoint in trials of first-line therapy for HER2-negative mBC, it has limitations. Designing trials to test potential new treatments for HER2-negative mBC is challenging and requires a balance of regulatory acceptability, feasibility, and realistic medical assumptions to calculate sample sizes, which can be particularly difficult in heterogeneous study populations with long postprogression survival and heterogeneous subsequent therapies. The magnitude of OS benefit likely to be considered as clinically (as well as statistically) significant depends on disease biology and risk. For example, in patients with triple-negative mBC, a 3-month improvement in median OS is undoubtedly meaningful, whereas in hormone receptor-positive mBC, a larger (6-month) improvement may be required to provide convincing meaningful benefit. In the current environment amid soaring costs and fierce competition, it is probably unrealistic to aim for trials demonstrating statistically significant OS improvement in this setting, except for trials in very specific poor prognosis populations. Ultimately, identification of robust alternative endpoints reflecting relevant patient benefits remains critical.

Data availability

All data used for the analyses reported in this paper are taken from the cited publications.

Acknowledgments

Medical writing support was provided by Jennifer Kelly, MA (Medi-Kelsey Ltd, Ashbourne, UK) and funded by Roche Pharma AG. This work was funded by Roche Pharma AG, Grenzach-Wyhlen, Germany. A poster reporting a preliminary analysis of this topic was presented at the European Society for Medical Oncology (ESMO) congress in Madrid, Spain; 8–12 September, 2017. The abstract from the poster was published in the congress proceedings.

Disclosure

SKÜ and CJ have participated on advisory boards for Roche Pharma AG. VM has participated on advisory boards for Amgen, AstraZeneca, Celgene, Daiichi Sankyo, Eisai, Genomic Health, Nektar, Novartis, Pfizer, Pierre Fabre, Roche, and Teva. SKI is an employee of Roche Pharma AG. MPL has participated on advisory boards for AstraZeneca, MSD, Novartis, Pfizer, Genomic Health, and Roche, and has received honoraria for lectures from Lilly, Roche, Novartis, Pfizer, Genomic Health, AstraZeneca, Medac, and Eisai. AS has declared no conflict of interest and the other authors report no other conflicts of interest in this work.

References

1. Mendes D, Alves C, Afonso N, et al. The benefit of HER2-targeted therapies on overall survival of patients with metastatic HER2-positive breast cancer – a systematic review. Breast Cancer Res. 2015;17:140.
2. Di Leo A, Bleiberg H, Buyse M. Overall survival is not a realistic end point for clinical trials of new drugs in advanced solid tumors: a critical assessment based on recently reported phase III trials in colorectal and breast cancer. J Clin Oncol. 2003;21(10):2045–2047.
3. Sargent DJ, Hayes DF. Assessing the measure of a new drug: is survival the only thing that matters? J Clin Oncol. 2008;26(12):1922–1923.
4. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst. 2009;101(23):1642–1649.
5. Amir E, Seruga B, Kwong R, Tannock IF, Ocaña A. Poor correlation between progression-free and overall survival in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies: a trial-level meta-analysis. Breast Cancer Res. 2015;17:140.
6. Liu L, Chen F, Zhao J, Yu H. Correlation between overall survival and other endpoints in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies: a trial-level meta-analysis. Breast Cancer Res Treat. 2015;154(3):591–608.
7. Adunlin G, Cyrus JW, Dranitis G. Correlation between progression-free survival and overall survival in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies: a literature-based analysis of 24 randomized trials. Bull Cancer. 2016;103(4):336–344.
8. Forsythe A, Chandiwana D, Barth J, Thabane M, Baekc j, Tremblay G. Progression-free survival/time to progression as a potential surrogate for overall survival in HR+, HER2- metastatic breast cancer. Breast Cancer. 2018;10:69–78.
## Supplementary material

### Table S1 Search strategy for identification of eligible trials

| Set# | Searched for                                                                 | Results |
|------|-----------------------------------------------------------------------------|---------|
| S1   | MESH.EXACT.EXPLODE("Breast Neoplasms")                                      | 260,430 |
| S2   | ti.ab((breast OR mamma*) NEAR/2 (cancer* OR carcinoma* OR tumo* OR neoplasm* or neoplasm* or malignanc*)) AND dstat.exact("Publisher" OR "In Process" OR "PubMed not MEDLINE" OR "In Data Review") | 32,411  |
| S3   | s1 OR s2                                                                    | 292,808 |
| S4   | ti.ab(metasta* or mBC or dissemin* or spread or advanced) AND s3              | 66,764  |
| S5   | MESH.EXACT.EXPLODE("Neoplasm Metastasis") AND s3                             | 31,899  |
| S6   | s4 OR s5                                                                    | 78,519  |
| S7   | s6 AND rtype.exact("Randomized Controlled Trial")                           | 2,832   |
| S8   | all(randomized) AND s6                                                       | 5,803   |
| S9   | all(placebo) AND s6                                                         | 475     |
| S10  | s7 OR s8 OR s9                                                              | 5,920   |
| S11  | (s7 OR s8 OR s9) AND rtype.exact("Clinical Trial, Phase III" OR "Clinical Trial, Phase II") | 969     |
| S12  | (ti.ab(phase p/2 III[*1] OR phase p/2 III[*2] OR phase p/2 II[*4] OR phase p/2 II[*4]) AND s10) | 1,468   |
| S13  | (s12) AND dstat.exact("Publisher" OR "In Process" OR "PubMed not MEDLINE" OR "In Data Review") | 126     |
| S14  | (s11 OR s13)                                                                | 1,095   |
| S15  | (s11 OR s13)) AND pd(2014-2018)                                             | 348     |