Physicians’ Perception of the Evidence in Relation to Primary Endpoints of Clinical Trials on Breast Cancer

Yi Zhang     Miao Liu     Houpu Yang     Shu Wang
Breast Center, Peking University People’s Hospital, Beijing, China

Keywords
Breast cancer · Primary endpoint · Clinical trial

Abstract

Objective: To investigate physicians' perception of the evidence of clinical trials on breast cancer. Methods: A survey was conducted by the Chinese Society of Breast Surgeons. We investigated the physicians’ perception of meaningful endpoints, appropriate follow-up duration, and clinically acceptable benefit through online questionnaires. Results: Among 278 validated questionnaires, the majority of the questions had no consistent answer. For local treatment, 30.6, 28.8, and 28.4% of participants regarded locoregional recurrence (LRR), disease-free survival (DFS), and overall survival (OS) as the most meaningful endpoint, respectively, and 47.5% believed that 5-year follow-up can alter clinical practice, and 34.5% thought it should be >10 years. In the adjuvant setting, 45.7, 38.5, and 12.9% regarded DFS, OS, and LRR as the most meaningful endpoint, respectively, and 52.5% thought that 10-year follow-up was solid, while 37.4% thought that 5-year follow-up was enough. In the advanced setting, 49.6, 24.1, and 23.7% considered progression-free survival, quality of life, and OS the most meaningful endpoint, respectively, and 39.6 and 28.8% considered that a follow-up of 1 year and 3 years, respectively, was meaningful. Similarly, the clinically acceptable absolute difference was inconsistent. Conclusion: Most Chinese oncologists advocated that surrogate endpoints could be used in certain circumstances, though OS was the most reliable one in breast cancer studies. Doctors’ perceptions of follow-up time and magnitude of benefit vary widely, reflecting the fact that there are many unanswered questions about supporting the use of new cancer treatments; a common understanding needs to be reached, such as a very consensual surrogate endpoint and a meaningful sufficiently large therapeutic benefit.

Introduction

Breast cancer is the most frequently diagnosed cancer among females worldwide and the leading cause of cancer deaths in females [1]. However, the death rate of female breast cancer patients has dropped by 38% from 1989 to 2014, which was partly attributed to the improvement of treatment methods [2]. Clinical trials, which assess the efficacy and toxicity of the new drugs, are the major approach to transform new breast cancer treatments into clinical practice. In the past few decades, the quantity and quality of clinical trials have improved significantly [3]. After a long history of scientific inquiry about the prevention, diagnosis, and treatment of human disease including breast cancer, we entered the era of evidence-based medicine, the core of which refers to an explicit and judicious use of current best evidence which was generated from clinical studies [4]. Evidence-based guidelines, mostly derived from appropriate evaluation of results of clinical studies, such as the National Comprehensive Cancer Network [5] and the European Society for Medical Oncology (ESMO) [6], have great impact on clinical decision-making worldwide and make clinical practice more standardized. To properly assess the strength of clinical studies, some
Physicians' Perception of Clinical Trials on Breast Cancer

methods to evaluate the level of evidence have been developed, such as the Oxford Centre for Evidence-Based Medicine levels of evidence [7, 8] and the Grading of Recommendations Assessment, Development and Evaluation system [9, 10]. These evidence level appraisal methods are mainly depended on the research type and the statistical outcome of the trials. For example, randomized trials are considered to have higher level of evidence in studies investigating treatment benefits. The outcome of studies was normally ranked as statistically positive or negative based on their statistical hypothesis and corresponding hypothesis testing. However, the real meaning of statistical benefit of a specific treatment to patients has not been considered enough in the current evidence hierarchy.

To help physicians accurately assess the value of a new therapy, the ESMO proposed a validated and reproducible approach, the ESMO Magnitude of Clinical Benefit Scale (ESMO-MCBS), which uses hazard ratio (HR) value and absolute gain in treatment outcomes to evaluate the magnitude of clinical benefit [11, 12]. Besides, the American Society of Clinical Oncology (ASCO) also brought forward a reference standard to help future clinical trials produce results that are clinically meaningful to patients [13]. Validations have been conducted in both early and advanced breast cancer, and they revealed that benefit assessments are robust grading tools to evaluate the benefit magnitude [14, 15].

However, these tools were not fully incorporated into the current evidence-based medicine system, and the real-world application of these frameworks is not really universal at present. It is still unclear how physicians think of the benefit of clinical trials in clinical practice. Therefore, it is necessary to investigate physicians' perception of the evidence of clinical trials on breast cancer. In the current study, we conducted a questionnaire-based survey in Chinese oncologists to explore attitudes regarding the endpoint choice, follow-up duration, and meaningful magnitude of clinical benefit in clinical trials on breast cancer.

**Methods**

**Survey Design**

The questionnaire included two sections. The first section of the questionnaire aimed to collect basic demographic information of participants, including sex, education background, professional title, practice duration, grade of the hospital, department, and specialty. The second section was designed to explore physicians' perception of evidence of clinical trials, including three parts: endpoints (overall survival [OS], defined as the time from randomization until death from any cause [16]; disease-free survival [DFS], defined as the time from randomization until disease recurrence or death from any cause [16]; progression-free survival [PFS], defined as the time from randomization until objective tumor progression or death [16]; locoregional recurrence [LRR]; breast cancer-specific survival; adverse effects; quality of life), follow-up duration, and magnitude of benefit. The following questions were included in this section: What is the most valuable endpoint for local therapy, adjuvant systemic therapy, and salvage therapy? How long do we need to follow up to get reliable results? What is the clinically acceptable absolute difference according to different primary endpoints? If more than half of the physicians gave the same answer, we considered having reached consensus (see online suppl. material for the complete questionnaire; for all online suppl. material, see www.karger.com/doi/10.1159/000518260). We did

| Table 1. Participant characteristics |
|------------------------------------|
| Question                          | n   | %    |
| Grade of hospital                  |     |      |
| Third-grade class A hospital       | 218 | 78.4 |
| Third-grade class B hospital       | 21  | 7.6  |
| Second-grade class A hospital      | 34  | 12.2 |
| Second-grade class B hospital      | 4   | 1.4  |
| Private hospital                   | 1   | 0.4  |
| Education background               |     |      |
| Doctor and master                  | 192 | 69.1 |
| Bachelor                           | 80  | 28.8 |
| Junior college                     | 6   | 2.2  |
| Professional title                 |     |      |
| Senior doctor                      | 97  | 34.9 |
| Associate senior doctor            | 93  | 33.5 |
| Attending                          | 62  | 22.3 |
| Resident                           | 16  | 5.8  |
| Researcher                         | 2   | 0.7  |
| Others                             | 8   | 2.9  |
| Department                         |     |      |
| Breast surgery                     | 160 | 57.6 |
| Breast medical oncology            | 15  | 5.4  |
| Surgical oncology                  | 6   | 2.2  |
| Medical oncology                   | 32  | 11.5 |
| General surgery (breast group)     | 26  | 9.4  |
| General surgery (ungrouped)       | 19  | 6.8  |
| Surgery                            | 3   | 1.1  |
| Others                             | 17  | 6.1  |
| Specialty                          |     |      |
| Only surgery                       | 16  | 5.8  |
| Surgery and adjuvant therapy       | 117 | 42.1 |
| Only medical oncology              | 45  | 16.2 |
| Both surgery and medical oncology  | 77  | 27.7 |
| Only basic research                | 1   | 0.4  |
| Only clinical research             | 8   | 2.9  |
| Only radiotherapy                  | 6   | 2.2  |
| Others                             | 8   | 2.9  |
| Years of practice                  |     |      |
| 1–5                                | 47  | 16.9 |
| 6–10                               | 68  | 24.5 |
| 11–20                              | 96  | 34.5 |
| >20                                | 67  | 24.1 |
| Geographic origin                  |     |      |
| Northern China                     | 147 | 52.9 |
| Southern China                     | 119 | 42.8 |
| Northwestern China                 | 8   | 2.9  |
| Tibetan regions                    | 2   | 0.7  |
| Overseas                           | 2   | 0.7  |

Breast Care 2022;17:180–187
DOI: 10.1159/000518260
not indicate a specific patient population, but just specified the questions from a macroscopic perspective, using adjuvant, neoadjuvant, and salvage settings.

**Survey Implementation**

The questionnaire was adapted to an electronic version and imported to an online open survey platform. The online survey was distributed at the annual symposium of the Chinese Society of Breast Surgeons (CSBS) held in May 2018. All the attendees of the CSBS symposium were invited to complete the questionnaire.

**Statistical Analyses**

The SPSS 25 software was used for statistical analyses. Descriptive statistics were used to analyze the survey results. Each question was plotted into a histogram for straightforward viewing (see online suppl. material for the histograms).

**Results**

**Participant Characteristics**

Among the total of 288 questionnaires we received, 278 were validated and 10 were ineligible (incomplete information or responders did not work in hospital). The participant characteristics are listed in Table 1. The majority \((n = 218, 78.4\%)\) of the participants practiced in third-grade class A hospitals, which is the highest level for hospitals in China. A total of 192 participants \((69.1\%)\) had a doctor or master degree. Ninety-seven \((34.9\%)\) participants had acquired the professional title of senior doctor, and 93 \((33.5\%)\) were associate senior doctors. The majority \((n = 160, 57.6\%)\) of the participants practiced in breast surgery departments. Years of practice varied: 47 \((16.9\%)\) had practiced for 1–5 years, 68 \((24.5\%)\) for 6–10 years, 96 \((34.5\%)\) for 11–20 years, and 67 \((24.1\%)\) for >20 years.

**The Most Valuable Endpoint**

Eighty-five \((30.6\%)\) participants regarded LRR as the most valuable endpoint in local treatment, and 80 \((28.8\%)\) and 79 \((28.4\%)\) considered it was DFS and OS, respectively (Fig. 1A). Only 31 physicians considered breast cancer-specific survival the most valuable endpoint. One \((0.4\%)\) participant pointed out that OS should always be the most important endpoint, but DFS is more meaningful when taking follow-up time into consideration. Nearly half of the participants \((n = 127, 45.7\%)\) regarded DFS as the most valuable endpoint in adjuvant systemic treatment in early-stage breast cancer (Fig. 1B), while 107 \((38.5\%)\) and 36 \((12.9\%)\) considered it was OS and LRR, respectively. In the systemic salvage treatment of advanced breast cancer, 138 \((49.6\%)\), 67 \((24.1\%)\), and 66 \((23.7\%)\) considered PFS, quality of life, and OS as the most valuable endpoint, respectively (Fig. 1C).

**Follow-Up Time**

A total of 132 \((47.5\%)\) participants believed that the results of a local treatment investigation with 5 years of follow-up can alter clinical practice, and 96 \((34.5\%)\) believed that the follow-up should be >10 years (Fig. 2A). When it comes to the investigation of adjuvant systemic treatment, 146 \((52.5\%)\) of the participants thought that 10 years of follow-up was solid to alter clinical practice, while 104 \((37.4\%)\) thought that 5 years of follow-up was enough (Fig. 2B). In the investigation of systemic salvage treatment, 110 \((39.6\%)\) and 80 \((28.8\%)\) participants considered a follow-up of 1 year and 3 years to be enough, respectively (Fig. 2C).

**Survival Benefit**

If the primary endpoint of a clinical trial was OS or DFS, more than one-third of the participants \((94 [33.8\%] \text{ and } 113 [40.6\%], \text{ respectively})\) believed that an absolute survival benefit of 5–10% would be meaningful in clinical practice (Fig. 3A, B). If the primary endpoint was LR, 102 \((36.7\%)\) thought that 5–10% would be meaningful...
(Fig. 3C). If the primary endpoint was PFS, 107 (38.5%) believed that an absolute survival benefit of 6–12 months would have significance (Fig. 3D).

**Discussion**

In an era of evidence-based medicine, well-designed clinical trials have a great impact on the updating of guidelines and clinical practice [4, 17]. However, the design of study endpoints and the definition of benefit magnitude were mixed, ranging from a few weeks of PFS to a few years of OS, which aroused concerns and controversies in clinical applications. Therefore, what constitutes meaningful clinical benefit is worth discussing, and understanding doctors’ attitudes toward clinical benefit is very important for future clinical research design and policy formulation. Our survey was conducted in this context. The results show that it is quite inconsistent in clinician opinion of study endpoint selection and benefit
magnitude, reflecting the fact that there are many unanswered questions about supporting the use of new cancer treatments; a common understanding needs to be reached.

As novel and expensive cancer treatments are continuously emerging, there has been increasing demand for assessment methods of the real clinical meaning to balance the benefit, cost, and patients’ expectations. The ASCO [13] and the ESMO [11, 12] have both proposed relatively specific approaches to evaluate the clinical benefit of new therapies. These tools were validated in different types of cancers [18]. A study applying the ESMO-MCBS to 59 individual trials and 6 meta-analyses demonstrated that the ESMO-MCBS is a robust tool for the evaluation of the magnitude of benefit studies in early breast cancer [15]. However, the ESMO-MCBS used the same method to evaluate clinical trials on different types of cancers, overlooking the fact that the development and molecular biological characteristics of different tumors vary. On the contrary, the ASCO framework just focused on specific scenarios. For breast cancer, only advanced triple-negative subtype was discussed. Therefore, the application of these frameworks is limited.

A meaningful clinical benefit of a trial refers to the endpoint, the follow-up time of the trial, and the magnitude of the benefit [19]. The selection of endpoint is vital in clinical trial design and approval of new drugs. OS is still the gold standard in clinical cancer trials [20]. However, it requires a larger number of patients and longer follow-up time, and may be confounded by subsequent treatment. Some surrogates, such as DFS and pathological complete response (pCR) in the (neo)adjuvant setting and PFS in advanced setting, are often designated as primary endpoints in cancer treatment trials in order to shorten the time required for drug development. However, the strength of the correlation between surrogate endpoints and OS is controversial [21].

DFS is a generally accepted surrogate for OS in adjuvant trials on breast cancer and was used by the Food and Drug Administration to support the accelerated approval of new drugs [22]. A study containing 126 trials showed that for a 1% increase in 2-year DFS difference, the 5-year OS difference increased by 0.5–0.55%; there was a statistically significant correlation between DFS and OS, although the correlation was not thought by the investigators to be strong enough for DFS to predict OS [23]. It is reasonable to think that an unclassified population may dilute the power of the correlation, and it has been validated that, in some specific molecular type of cancer, the correlation between DFS and OS was stronger, like in the HERA trial, with the addition of trastuzumab to chemotherapy compared with chemotherapy alone in human epidermal growth factor receptor 2 (HER2)-positive breast cancer, where an absolute benefit in terms of DFS at 2 years of 8.4% was reported [24], while conferring an absolute benefit of 6.8% improvement in 10-year DFS and a 6.5% absolute gain in OS at 12 years [25]. The results of the HERA trial indicate that DFS may shorten the follow-up time through producing positive results earlier. In a systemic review and meta-analysis, eight trials (n = 21,480 patients) were included in the analysis; patient-level associations between DFS and OS were strong (r = 0.90, 95% CI 0.89–0.90) [26]. Further meta-analyses should be conducted to explore the value of DFS in luminal and triple-negative subtypes. In ESMO-MCBS, both OS and DFS were selected for benefit evaluation in curative intent trials. In our study, 38.5% of the participants believed that OS remained the most meaningful endpoint, and 45.7% considered that DFS can also be an appropriate endpoint in early-stage trials. DFS is the most commonly used surrogate endpoint for OS in adjuvant trials. Three years was defined as the minimum appropriate follow-up time in ESMO-MCBS. In our study, all physicians agreed that 3 years of follow-up was meaningful, but actually 94.2% expected >5 years of follow-up, and 56.8% preferred ≥10 years. The requirement of follow-up time was stricter in our study.

pCR is another commonly used surrogate endpoint in many neoadjuvant studies [27], but whether pCR could be a reliable endpoint is quite controversial. Some studies demonstrated pCR after neoadjuvant therapy as a prognostic indicator for OS [28, 29], and it can be used as a valid surrogate endpoint [27, 30]. However, there are studies coming to an opposite conclusion [31]. In the National Surgical Adjuvant Breast Project B-27 (NSABP B-27) trial, taxanes in addition to anthracycline and cyclophosphamide achieved a doubled pCR rate increase from 13 to 26%, but only a 1.3% (from 75.5% to 76.8%) OS benefit, without statistical significance [32]. A pooled analysis proved that the association between pCR and long-term outcomes was strongest in patients with triple-negative and HER2-positive, hormone receptor-negative tumors, but trial-level analysis showed little association between increases in pCR rate and OS [33]. Recently many studies have considered that the pCR rate is quite different in various intrinsic subtypes of breast cancer, and the magnitude of pCR improvement translating into long-term survival is also different in varying subtypes [28]. Even in the same subtype but with different regimens, a similar increase in pCR does not result in similar survival improvement, like the NeoALTO [34, 35] and ALTTO [36] trials versus the NeoSphere [37, 38] and APHINITY [39] trials. The NeoALTO trial [34, 35] showed a significant increase in pCR rate in the neoadjuvant setting, from 29.5% in the trastuzumab group to 51.3% in the lapatinib plus trastuzumab group, but the ALTTO [36] trial, which use the same regimen in the adjuvant setting, did not support the significant DFS improvement. Similarly in the NeoSphere trial [37, 38], pCR
physicians were decreased consequently. We support the ASCO’s advocacy that designs research for a specific clinical problem to maximize clinical benefit [13]. Defining a smaller benefit goal will consume more patient resources and push the trial design to a negative circle – smaller and smaller gain with more and more patients needed. The ASCO advocates that funders of clinical studies fully understand the perspectives of patients and physicians, weigh the pros and cons, and set clear endpoints, raising the bar for clinical trials by defining clinically meaningful outcomes. Also, the working group thought that incremental gains should be accompanied by little to no increase in toxicity if the new therapies had more toxic than current standards; the greatest increments in OS should also be produced [13].

The meaningful magnitude of clinical benefit changes with time. Contemporary breast cancer treatment was on the shoulders of giants, on which dramatic benefits were less likely to be obtained through traditionally designed trials. In the era of Halsted, radical mastectomy could increase OS from 20 to 42.3% [46, 47], which has never been seen thereafter. In the NSABP B-05 trial [48], L-phenylalanine mustard prolonged DFS by 12.3% in comparison with the control arm. Due to significant improvement in early detection and comprehensive treatment, it seems more and more difficult to attain a large survival benefit. In the aspect of target therapy in HER2-positive breast cancer, the HERA trial [49] achieved a 4-year DFS improvement of 6.4% due to the addition of 1 year of trastuzumab; however, in the APHINITY trial [39], dual blockade of trastuzumab and pertuzumab just achieved 1.7% invasive disease-free survival (iDFS) improvement in comparison with single-target therapy. Study-designed focusing on higher-risk patients helped to magnify the benefit. In recent years, APHINITY and KATHERINE have been two vital clinical trials in the field of breast cancer and have had a marked impact on clinical decision-making in adjuvant therapy. Although both trials achieved a statistical difference, the absolute benefit varies. The APHINITY trial [39] is a randomized controlled trial which investigated whether the addition of pertuzumab to trastuzumab can improve outcomes versus trastuzumab plus placebo among patients with HER2-positive breast cancer in adjuvant treatment. The 4-year iDFS of the pertuzumab group was 92.3%, while that of the placebo group was 90.6% (HR = 0.81, p = 0.045). In the subgroup of node-positive patients, the 4-year iDFS was 89.9% in the pertuzumab group and 86.7% in the placebo group (HR = 0.77, p = 0.02). The KATHERINE trial [50] is a multicenter phase 3 randomized trial which aimed to investigate the effectiveness of T-DM1 in breast cancer patients who have residual invasive disease after neoadjuvant therapy. Patients received either T-DM1 or trastuzumab. The 3-year iDFS of the T-DM1 group was 88.3%, which was significantly higher than that of the trastuzumab group (77.0%) (HR = 0.50, p < 0.001). In the APHINITY trial, iDFS was increased by 1.7% in the pertuzumab group in comparison with the placebo group, and the minor difference has raised debates whether we should use expensive dual blockade to each HER2-positive patient. Meanwhile, the KATHERINE trial achieved 11% absolute benefit when using T-DM1 in non-pCR patients, which resulted in undisputed consensus. The wiser study design and the explora-
tion in a selected population with higher relapse risk according to the status of non-pCR contributed to the most part to the much larger survival benefit of KATHERINE. Smaller and smarter trials need to be designed by picking up specific populations by pathological and molecular characteristics and using specific therapy; the design of a trial is increasingly important in the time of target and immune therapy.

There are some limitations to our study. The purpose of investigating new treatments is to provide benefit to patients. However, drugs usually bring unexpected adverse effects and extra economic burden. Our study did not take the patients’ ideas into consideration. How patients think of a therapy has been explored for a long time [51, 52]. A previous study reported that for most breast cancer patients, 1-year additional survival is worthwhile receiving 6 months of chemotherapy [52]. Different from the active attitude to treatment of patients, physicians tend to be more conservative. When benefit is limited, physicians are less likely to recommend chemotherapy [53]. Physicians and patients ought to reach an agreement on this issue; the therapeutic regimen should be recommended by physicians and thoroughly discussed with the patients.

Conclusion

Our survey result showed that doctors’ perceptions of clinical evidence vary widely; common understanding needs to be reached such as a very clear surrogate endpoint and meaningful sufficiently large therapeutic benefit, which is very important in transformation from trial to clinical use. In order to achieve clinical significance, researchers should lay more stress on trial design, especially the selection of endpoints and enrolled patients. Patients’ perspective and expectations should also be considered; further studies should be conducted.

Statement of Ethics

Consent was indicated in the introduction of the questionnaire and obtained when the participants chose to submit the survey (see online suppl. material).

Conflict of Interest Statement

The authors declare no competing nonfinancial/financial interests.

Funding Sources

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions

Y. Zhang collected and analyzed the data and wrote the manuscript. M. Liu conducted the survey and analyzed the data. S. Wang designed the questionnaire and analyzed the data. H. Yang designed the questionnaire and revised the manuscript.

References

1 Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021 May;71(3):209–49.
2 Siegel RL, Miller KD, Jemal A. Cancer statistics. 2017. CA Cancer J Clin. 2017;67(1):7–30.
3 Zhou Q, Chen XY, Yang ZM, Wu YL. The changing landscape of clinical trial and approval processes in China. Nat Rev Clin Oncol. 2017 Sep;14(9):577–83.
4 Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn’t. BMJ; 1996; 312(7023):71–2.
5 National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology. Available from: https://www.nccn.org.
6 European Society for Medical Oncology (ESMO). ESMO Guidelines. Available from: https://www.esmo.org/guidelines.
7 OCEBM Levels of Evidence Working Group. The Oxford levels of evidence 1. Oxford Centre for Evidence-Based Medicine. Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/combined-levels-of-evidence.
8 OCEBM Levels of Evidence Working Group. The Oxford levels of evidence 2. Oxford Centre for Evidence-Based Medicine. Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence.
9 Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ; 2008;336(7650):924–6.
10 Guyatt GH, Oxman AD, Akil EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011 Apr;64(4):383–94.
11 Cherny NI, Sullivan R, Dafni U, Kerst JM, Sobrero A, Zielinski C, et al. Application of the ESMO-Magnitude of Clinical Benefit Scale (V.1.1) to the field of early breast cancer therapies. ESLO Open. 2020 Sep;5(5):e000743.
Physicians’ Perception of Clinical Trials on Breast Cancer