Contemporary Clinical Trials 53 (2017) 20–28

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Surrogacy of progression free survival for overall survival in metastatic breast cancer studies: Meta-analyses of published studies

Madan G. Kundu*, Suddhasatta Acharyya

Novartis Pharmaceutical Corporation, East Hanover, NJ, USA

ARTICLE INFO

Article history:
Received 30 July 2016
Received in revised form 30 November 2016
Accepted 3 December 2016
Available online 8 December 2016

Keywords:
Metastatic breast cancer
Progression free survival
Overall survival
Surrogacy
Meta-analysis
ROC curve
Classification tree
Surrogate threshold effect

ABSTRACT

Purpose: PFS is often used as a surrogate endpoint for OS in metastatic breast cancer studies. We have evaluated the association of treatment effect on PFS with significant HR_{OS} (and how this association is affected by other factors) in published prospective metastatic breast cancer studies.
Methods: A systematic literature search in PubMed identified prospective metastatic breast cancer studies. Treatment effects on PFS were determined using hazard ratio (HR_{PFS}), increase in median PFS (ΔMED_{PFS}) and % increase in median PFS ($\%\Delta MED_{PFS}$). Diagnostic accuracy of PFS measures (HR_{PFS}, ΔMED_{PFS} and $\%\Delta MED_{PFS}$) in predicting significant HR_{OS} was assessed using receiver operating characteristic (ROC) curves and classification tree approach (CART).
Results: Seventy-four cases (i.e., treatment to control comparisons) from 65 individual publications were identified for the analyses. Of these, 16 cases reported significant treatment effect on HR_{OS} at 5% level of significance. Median number of deaths reported in these cases were 153. Area under the ROC curve (AUC) for diagnostic measures as HR_{PFS}, ΔMED_{PFS} and $\%\Delta MED_{PFS}$ were 0.69, 0.70 and 0.75, respectively. Classification tree results identified $\%\Delta MED_{PFS}$ and number of deaths as diagnostic measure for significant HR_{OS}. Only 7.9% (3/39) cases with ΔMED_{PFS} shorter than 48.27% reported significant HR_{OS}. There were 7 cases with ΔMED_{PFS} of 48.27% or more and number of deaths reported as 227 or more – of these 5 cases reported significant HR_{OS}.
Conclusion: $\%\Delta MED_{PFS}$ was found to be a better diagnostic measure for predicting significant HR_{OS}. Our analysis results also suggest that consideration of total number of deaths may further improve its diagnostic performance. Based on our study results, the studies with 50% improvement in median PFS are more likely to produce significant HR_{OS} if the total number of OS events at the time of analysis is 227 or more.

1. Introduction

As per national cancer institute, in the U.S., breast cancer is the second most common non-skin cancer and the second leading cause of cancer-related deaths in women; and, therefore, there has always been a high demand for novel breast cancer therapies. At the time of preparing this manuscript, based on ClinicalTrial.gov search [1], 175 phase III breast cancer studies were actively recruiting patients. For breast cancer therapies, the main goal is to improve overall survival (OS) and quality of life [2,3]. US FDA guideline [4] states that "[overall] survival is considered the most reliable cancer endpoint". Due to the advancement in metastatic breast cancer management and therapies, there has been marked improvement in OS in breast cancer patients in the last few decades. Consequently, patients need to be followed-up for longer period of time to observe sufficient number of OS events (i.e., deaths) [5] before treatment effect on OS can be evaluated statistically. Further, as many patients switch to second line (and beyond) therapies upon progression, the OS time may be influenced by post-progression therapy. For these reasons, surrogate endpoints such as progression-free survival (PFS) or objective response rate (ORR) are being increasingly used for accelerated approvals, with PFS being the one used most often [2]. The basis for using PFS as surrogate endpoint for OS is as follows: cancer progression represents an ominous march toward death from malignancy. Hence, the longer it takes for the cancer to progress, the longer a patient will live [6]. In general, PFS has not been statistically validated for surrogacy of OS yet in breast cancer studies [2,4]. Reported results regarding association between Hazard ratio of PFS and OS in the metastatic breast cancer studies are mixed: For example, Hackshaw et al. [7] found a correlation of 0.87; Burzykowski et al. [8] reported correlation of 0.48; Michiels et al. [9] reported R^2 (i.e. proportion of the variance in the true endpoint that is explained by the surrogate endpoint) as 0.51.

According to Prentice's definition [10], in order for PFS to be a "statistically validated" surrogate endpoint for OS, "test for null

* Corresponding author at: 1 Health Plaza, East Hanover, NJ 07936, United States.
E-mail address: madan_g.kundu@novartis.com (M. Kundu).

http://dx.doi.org/10.1016/j.cct.2016.12.004
1551-7144/

hypothesis of no treatment effect in PFS" should be a valid "test for null hypothesis of no treatment effect in OS". The test for treatment effect on OS is carried out by testing $HR_{OS}=1$, where HR_{OS} is the hazard ratio (HR) of OS. However, many randomized clinical trials failed to demonstrate significant treatment effect in OS despite demonstrating significant treatment effect in PFS. The current project attempts to investigate the trial level surrogacy in breast cancer studies from a diagnostic testing perspective using nonparametric approaches. It is important to note that our investigation differs from previous investigations [7,8,11-16] based on meta-analytic methods, where the primary purpose was to examine the strength of treatment effect on PFS to predict treatment effect on OS at trial level. The definition of trial level surrogacy in the current investigation is intuitive and aligned with the ultimate question that all stakeholders, regulators in particular, are often seeking an answer to, from a phase III cancer clinical trial – Is there a statistically significant OS benefit in the new treatment that is discernible from the data on progression-free survival (PFS) in metastatic breast cancer studies? This definition of trial level surrogacy was also considered by Burzykowski and Buyse [17] as it can be useful to estimate the "Surrogate threshold effect". Surrogate threshold effect can be defined as the minimum treatment effect on PFS measure that is required to predict statistically significant HR_{OS}.

Our goal was to evaluate the trial level surrogacy of PFS for OS solely based on published clinical trial results. Burzykowski et al. [8] evaluated trial level surrogacy by fitting simple (log-) linear regression analysis to model HR_{OS} with ratio of median PFS time and then used R^2 as a measure of trial level surrogacy. Buyse et al. [18] proposed to estimate trial level surrogacy using R^2 as well, but in a more sophisticated way using trial specific random effects. These methods make various model assumptions such as PFS and OS are linearly associated [8] or some distributional assumption [18]. As Venook and Tabernero [6] have pointed out association of PFS with OS may be complicated in today's era and, therefore, a simplified linear model may not be sufficient to describe the association. Further, the use of R^2 is heavily impacted by the presence of outlier [19]. Another problem related to R^2 is the difficulty in interpreting its value [17]. For these reasons, we have adopted non-parametric approaches to evaluate the trial level surrogacy which, unlike parametric methods, do not require to make distributional assumptions or to pre-specify the from of the association. The advantages of non-parametric methods are that these methods are completely data-driven and free from model assumptions. Consequently, non-parametric methods have obvious advantage of producing results which are solely based on observed data and are not dependent on unverifiable model assumptions. Non-parametric methods can be also useful (a) to find out which PFS measure is relatively more important in predicting significant HR_{OS}, (b) to study the influence of other factors (e.g., sample size and total number of events) on the association of PFS measure and significant HR_{OS} as, for example, the power for statistical test of HR_{OS} is a function of total number of OS events, and (c) to estimate surrogate threshold effect. Results from non-parametric methods are often easy to interpret, and allow granular visualization of the results. For this project, breast cancer studies were our focus, but the similar investigation can be carried out for other indications as well. Throughout the article, (unless otherwise mentioned), 'statistically significant' would imply that the significance was in favor of the treatment.

2. Methods

2.1. Literature search

A systematic literature search in PubMed (July 2015) was performed to identify published prospective studies on metastatic breast cancer research with both PFS and OS comparison results reported. The search syntax used was as follows: "(((Breast Cancer[Title]) AND Randomized[Title/Abstract]) AND Progression free survival[Text Word]) AND Overall survival[Text Word]". The PubMed search returned 181 publications between Jul-2000 and Jul-2015. Many of these studies were systematic literature review or meta-analyses and hence dropped. Further, studies with either PFS or OS not reported were also excluded. We were able to find 64 individual prospective studies [20-73,75-84] where both PFS and OS comparison results were reported. In addition, in one publication [74], instead of PFS, time to progression (TTP) was reported and that study was included. Therefore, we had total of 65 publications for the meta-analyses.

2.2. Data extraction

Of the 65 selected publications, in seven prospective studies [25-27,57,59,61,84], two pairs of treatment-to-control comparisons were reported and in one prospective study [35], three pairs of treatment-to-control comparisons were reported. Therefore, we had total of 74 treatment-to-control comparison available for the meta-analyses. For each treatment-to-control comparison, the following information were extracted: randomization status, blinding status (open or blinded), total sample size (treatment plus control), total number of events (treatment plus control), median PFS, median OS, HR (hazard ratio) in PFS (HR_{PFS}), HR in OS HR_{OS}, reported p-value (or significance status) for HR_{PFS} and reported p-value (or significance status) for HR_{OS}. In case both local and central PFS assessments were reported, the one which was reported as primary endpoint was considered.

2.3. Statistical methods

Treatment effect on PFS was determined using the following measures: hazard ratio (HR_{PFS}), increase in median PFS (ΔMED_{PFS}) and % increase in median PFS ($\%\Delta MED_{PFS}$). All three measures were used as diagnostic tools for predicting statistically significant HR_{OS} in favor of treatment (yes/no).

We have assessed the trial level surrogacy of PFS for OS by evaluating the diagnostic accuracy of these comparative PFS measures to predict statistically significant HR_{OS}. Diagnostic accuracy of comparative PFS measures (HR_{PFS}, ΔMED_{PFS} and $\%\Delta MED_{PFS}$) in predicting significant HR_{OS} was assessed using receiver-operating characteristic (ROC) curve [85], and classification tree (using CART algorithm [86]) approach. Empirical ROC curves were drawn plotting the true positive rate (proportion of correct prediction of significant HR_{OS} based on comparative PFS measure among those reporting significant treatment effect on HR_{OS}) against the false positive rate (proportion of wrong prediction of significant HR_{OS} based on PFS measure among those reported non-significant treatment effect on HR_{OS}). True positive rate, and false positive rate were obtained at each unique value of comparative PFS measures. For a given unique value of x, if comparative PFS measure was greater than or equal to x, then it was predicted that HR_{OS} will be significant; otherwise not. The accuracy of the diagnostic measure was assessed by numerically computing the area under ROC curve (AUC), with larger AUC implying better accuracy. Optimal cut-off points based on ROC curve were identified according to Youden's index [87]. According to Youden's criteria an optimum cut-off point for prediction of significant HR_{OS} would be one that maximizes the difference between true positive rate and false positive rate.

We have utilized classification tree to answer following questions: (a) which trial level measure of treatment benefit in PFS has stronger association with significant HR_{OS} in favor of treatment – HR_{PFS} or (%) median improvement in PFS? (b) Is there any other factor(s) (e.g., total number of deaths) that influence significance of HR_{OS}? (c) if yes, then how does this measure modify the association