Is response rate increment obtained by molecular targeted agents related to survival benefit in the phase III trials of advanced cancer?


1Department of Respiratory Medicine, Allergy and Rheumatic Diseases, Osaka University Graduate School of Medicine, Suita, Osaka; 2Department of Medical Physics, School of Health Sciences, Kumamoto University, Kumamoto; 3Department of Internal Medicine, National Hospital Organization Kinki-chuo Chest Medical Center, Sakai, Osaka; 4Department of Medical Oncology, Kinki University School of Medicine, Osaka; 5Division of Respiratory Medicine and Allergology, Department of Internal Medicine, Aichi Medical University School of Medicine, Aichi and 6Department of Medical Oncology, Sakai Hospital, Kinki University School of Medicine, Sakai, Osaka, Japan

Received 11 August 2009; revised 12 November 2009; accepted 19 November 2009

Background: It remains unclear whether response rate (RR) is related to survival benefit in phase III trials of advanced cancer treated with molecular targeted agents (MTA) in combination with standard therapies.

Materials and methods: We carried out a systematic search of PubMed for randomized phase III trials of four solid tumors examining the efficacy of MTA when added to a standard therapy. We examined whether there were any associations between RR increment obtained by the addition of targeted agents (ΔRR) and survival benefit in phase III trials.

Results: We identified 26 phase III trials of MTA with a total of 21 156 patients and 29 experimental arms of MTA. Studies which showed significant survival benefit had higher ΔRR compared with those which did not show significant benefit. In the receiver operating characteristic curve analysis, using a 7% gain as threshold value for ΔRR allowed assessment of survival benefit with high sensitivity and specificity. There were also significant relationships between ΔRR and hazard ratios for overall survival and progression-free survival in the linear regression analysis.

Conclusion: RR increment obtained by the addition of MTA to a standard therapy may be useful to predict survival benefit in clinical phase III trials of advanced cancer.

Key words: molecular targeted agents, response rate, survival benefit

introduction

In recent years, targeted therapies have become an important modality in cancer treatment. A growing number of new agents targeting molecular pathways are tested individually or in combination with standard treatments in clinical trials. Gaining higher response rate (RR) has usually been determined as a primary end point in phase II studies evaluating the efficacy of new cytotoxic drugs [1]. However, molecular targeted agents (MTA) appear to be different from cytotoxic agents in their tumor-killing mechanism, and therefore, RR might not be an appropriate end point for phase II or screening studies of MTA [2, 3]. Of course, demonstrating an association of RR and survival in phase III trials does not imply that RR can be used instead of survival for registrational trials. The latter use of a surrogate requires a stronger relationship such as was shown by Sargent et al. [4] and used in the decision to approve oxaliplatin for adjuvant colon cancer.

We have recently indicated that RR could be a surrogate marker for survival in the clinical trials of non-small-cell lung cancer treated with gefitinib or erlotinib [5]. However, as noted above, to enable registrational trials to use RR instead of survival would require a dataset indicating that a surrogate threshold effect (STE) was exceeded by the RR found in the particular registrational trials. El-Maraghi and Eisenhauer [6] reported that objective RR seemed to be a useful end point for new targeted agents because, in their review, its observation was predictive of regulatory approval for US Food and Drug Administration. However, their analyses were limited to trials of single agents, and the authors did not discuss if their conclusions were valid for combinations of MTA with cytotoxics or others.

It is not easy to predict the effectiveness of combination therapies from results of their single-agent studies [7]. Consequently, not only single-agent phase II studies but also combination phase II studies seem to be necessary to investigate whether there is enough evidence of benefit of the new combination to warrant a phase III combination trial. Therefore, it is crucial to identify appropriate end points of combination phase II studies.
The purpose of this study is to investigate through a systematic review of publications whether RR is useful to predict survival benefit in phase III trials of advanced cancer treated with MTA in combination with standard therapies. To our knowledge, this is the first analysis investigating the relationships between RR and survival benefit in phase III trials of MTA in combination with standard therapies. As many targeted agents continue to be used in combination with established chemotherapies, it is important to discuss this issue considering novel agents yet to be developed.

materials and methods

literature search and data extraction

We carried out a systematic search of PubMed for randomized phase III trials examining the efficacy of MTA when added to a standard therapy. We targeted randomized phase III trials evaluating survival benefit which compared a combination therapy of a standard therapy and MTA with the standard therapy alone or with a placebo. All trials that had been reported by 31 May 2009, were targeted. Biotherapy such as immune therapy or hormonal therapy was not considered as a molecular targeted therapy in our analyses. We defined four common solid cancer types (lung, colorectal, breast, and renal cell carcinoma) on which to focus.

Systematic search was carried out using the key words lung cancer, colorectal cancer, breast cancer, and renal cell carcinoma. All searches were limited to 'English language' and 'clinical trial, phase III' or 'randomized controlled trial'. Retrospective subgroup analyses of phase III trials were not included. Trials involving radiation therapy or using MTA as adjuvant, neoadjuvant, or consolidation therapies were excluded from our analysis.

For each trial, data on sample size, primary and secondary end points, RR, overall survival (OS), progression-free survival (PFS), time to progression (TTP), types of MTA, and therapy regimens were collected. Also investigated in each study was whether statistically significant benefit of OS or PFS was observed in an experimental arm of MTA compared with a control arm. When available, hazard ratios (HRs) for OS and PFS were collected. We also recorded data of study designs on response evaluation to assess heterogeneity across studies.

All phase III studies were reviewed independently by two investigators (KT and TT) to assess the reliability of data extraction.

Statistical analysis

For each trial, the difference in RR (ΔRR) was calculated as the estimate in the experimental arm minus the estimate in the control arm. Using Student’s t-test, we examined whether there are any differences in the distribution of ΔRR between trials which showed survival benefit and those which did not show survival benefit. In addition, we calculated the area under receiver operating characteristic (ROC) curves (AUCs) to examine the accuracy of ΔRR to predict survival benefit in phase III studies. A binormal ROC curve was fitted to ΔRR for positive and negative groups classified by the presence of significant benefit in OS or PFS. By use of the actual value of ΔRR, a computer program (ROCKIT program [8], kindly provided by Metz, The University of Chicago) was employed for estimating binormal ROC curves as well as the AUC value and its 95% confidence interval.

Linear regression analysis was carried out to evaluate the correlation between ΔRR and HR of OS or PFS. In this model, the prediction bands were calculated and the STE was determined by the intersection of the upper prediction band and the HR = 1 line. This method estimates the threshold level of a surrogate needed in a new individual clinical trial to predict gain in the target outcome [9]. Concerning the heterogeneity of the tumor types, linear regression analysis was also carried out limited to only lung cancer studies and colorectal cancer studies. Given the small number of studies for the other tumor types, insufficient data were available to carry out statistical analysis. Linear regression analysis was also employed to evaluate the correlation between HR of PFS and that of OS. A P value <0.05 was considered statistically significant, and all reported P values were two sided. All statistical analyses except those of ROC were carried out using SPSS 17.0 for Windows (SPSS, Inc., Chicago, IL).

results

study characteristics

We identified 26 phase III trials of MTA with a total of 21,156 patients and 29 experimental arms of MTA in combination with standard therapy. The baseline characteristics of the 26 trials are shown in Table 1 [10–35]. There were 10 lung cancer studies, 7 colon cancer studies, 6 breast cancer studies, and 3 renal cell carcinoma studies.

The sample size of each arm with MTA ranged from 228 to 1401, with a median sample size of 773 patients. Four trials included death in the definition of TTP, and therefore, they were calculated as PFS in our analysis. Twenty-five study arms with MTA contained mature data of PFS, and 12 of them reported statistically significant benefits of experimental arms (arms with MTA) compared with control arms. In contrast, 25 study arms with MTA contained mature data of OS, and 5 of them reported statistically significant benefits. RR was reported in all included studies, and ΔRR ranged from −7.9% to 20.0% (median: 5.5%).

A primary end point and data on response evaluation in each study are summarized in supplemental Table S1 (available at Annals of Oncology online). In the trials we used, the primary end point most frequently used was OS, followed by PFS. In addition, there appeared to be some differences among each study in the methods of response evaluation.

validity of regression model

For all regression analyses, none of the tests for normality of error and heteroscedasticity were statistically significant.

RR and survival benefit

Studies which showed significant benefit in OS had higher ΔRR compared with those which did not show significant benefit in OS (P = 0.002, Figure 1A). Similarly, studies with significant benefit in PFS had higher ΔRR compared with those without significant benefit in PFS (P < 0.001, Figure 1B). In an ROC analysis, the AUC value and its 95% confidence interval for predicting the presence of significant benefit in OS by ΔRR were 0.903 (0.670–0.985) and that for predicting the presence of significant benefit in PFS by ΔRR were 0.910 (0.730–0.985) (Figure 1C and D). Both performed significantly better than random guessing of survival benefit. Using a 7% gain as threshold value for ΔRR allowed assessment of survival benefit with a sensitivity of 91% and a specificity of 70% in OS and with a sensitivity of 84% and a specificity of 85% in PFS.

Similar findings were observed in subgroup analyses of individual cancer types. A 7% gain of ΔRR significantly correlated with the presence of survival benefit in phase III study in each cancer type (supplemental Table S2, available at Annals of Oncology online). Notably, all lung cancer studies
with ΔRR no less than 7% showed significant benefit in at least either of OS or of PFS, whereas no lung cancer study with ΔRR <7% showed significant benefit in any survival.

RR and survival improvement

Eighteen studies with 18 arms for OS and 17 studies with 18 arms for PFS had data of HRs. There were significant relationships between ARR and HR for OS or PFS on unweighted linear regression analysis (\( P = 0.002 \) and 0.001, \( R^2 = 0.47 \) and 0.50, respectively, Figure 2A and B). The STE levels of ΔRR calculated from prediction bands were 21% for OS and 15% for PFS. The regression analysis weighted for trial size showed similar results (data not shown).

Linear regression analyses were also carried out limited to lung or colorectal cancer studies (Figure 2). Despite the number of evaluable studies being small, ΔRR in lung cancer studies showed strong relationships with both HR for OS and that for PFS. Similarly, a significant relationship was observed between ΔRR and HR for PFS in colorectal cancer studies.

PFS and OS

Only 14 studies reported both HR of PFS and that of OS. In unweighted regression analysis, HR of PFS significantly correlated with that of OS (\( P < 0.001, R^2 = 0.69; \) supplemental Figure S1, available at Annals of Oncology online). The regression analysis weighted for trial size showed similar results (data not shown).

Discussion

Possible interactions in therapeutic effects of MTA and standard therapies have made it difficult to predict the relation between clinical response and survival benefit. Our analysis showed, on average, that the addition of MTA to standard

Table 1. Trials included in the analysis (\( N = 26 \))

<table>
<thead>
<tr>
<th>Cancer type/agents</th>
<th>Target</th>
<th>Drug type</th>
<th>Trials</th>
<th>Control arms</th>
<th>Placebo controlled</th>
<th>No. of patients</th>
<th>Primary end point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gefitinib</td>
<td>EGFR</td>
<td>SM</td>
<td>Giaccone et al., 2004 [10]</td>
<td>CDDP + GEM</td>
<td>No</td>
<td>1093</td>
<td>OS</td>
</tr>
<tr>
<td>Erlotinib</td>
<td>EGFR</td>
<td>SM</td>
<td>Herbst et al., 2004 [11]</td>
<td>CBDCA + PTX</td>
<td>No</td>
<td>1037</td>
<td>OS</td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>VEGF</td>
<td>AB</td>
<td>Gatzemeier et al., 2007 [13]</td>
<td>CBDCA + PTX</td>
<td>No</td>
<td>1079</td>
<td>OS</td>
</tr>
<tr>
<td>Cetuximab</td>
<td>EGFR</td>
<td>AB</td>
<td>Sandler et al., 2006 [14]</td>
<td>CDDP + GEM</td>
<td>No</td>
<td>1159</td>
<td>OS</td>
</tr>
<tr>
<td>Prinomastat</td>
<td>Matrix metalloproteinase</td>
<td>SM</td>
<td>Bissett et al., 2005 [17]</td>
<td>CBDPA + GEM</td>
<td>Yes</td>
<td>1125</td>
<td>OS</td>
</tr>
<tr>
<td>BMS-275291</td>
<td>Matrix metalloproteinase</td>
<td>SM</td>
<td>Leight et al., 2005 [18]</td>
<td>CBDCA + PTX</td>
<td>Yes</td>
<td>362</td>
<td>OS</td>
</tr>
<tr>
<td>Aprinocarsen</td>
<td>Protein kinase C alpha</td>
<td>AS</td>
<td>Paz-Ares et al., 2006 [19]</td>
<td>CDDP + GEM</td>
<td>No</td>
<td>774</td>
<td>OS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>VEGF</td>
<td>AB</td>
<td>Hurwitz et al., 2004 [20]</td>
<td>FOLFOX</td>
<td>Yes</td>
<td>813</td>
<td>OS</td>
</tr>
<tr>
<td>Panitumumab</td>
<td>EGFR</td>
<td>AB</td>
<td>Saltz et al., 2008 [22]</td>
<td>FOLFOX or XELOX</td>
<td>Yes</td>
<td>1401</td>
<td>PFS</td>
</tr>
<tr>
<td>Cetuximab</td>
<td>EGFR</td>
<td>AB</td>
<td>Hecht et al., 2009 [23]</td>
<td>ox-Ch or Iri-CT + Bev</td>
<td>No</td>
<td>1053</td>
<td>PFS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trastuzumab</td>
<td>HER-2</td>
<td>AB</td>
<td>Slamon et al., 2001 [27]</td>
<td>Standard chemotherapy*</td>
<td>No</td>
<td>469</td>
<td>TTP</td>
</tr>
<tr>
<td>Lapatinib</td>
<td>EGFR/HER-2</td>
<td>SM</td>
<td>Geyer et al., 2006 [29]</td>
<td>Capcitabine</td>
<td>No</td>
<td>324</td>
<td>TTP</td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>VEGF</td>
<td>AB</td>
<td>Miller et al., 2005 [31]</td>
<td>Capcitabine</td>
<td>No</td>
<td>462</td>
<td>PFS</td>
</tr>
<tr>
<td>Renal cell carcinoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bevacizumab</td>
<td>VEGF</td>
<td>AB</td>
<td>Escudier et al., 2007 [33]</td>
<td>Interferon alfa</td>
<td>Yes</td>
<td>649</td>
<td>OS</td>
</tr>
<tr>
<td>Temsirolimus</td>
<td>mTOR</td>
<td>SM</td>
<td>Hudes et al., 2007 [35]</td>
<td>Interferon alfa</td>
<td>No</td>
<td>732</td>
<td>OS</td>
</tr>
</tbody>
</table>

EGFR, epidermal growth factor receptor; SM, small molecule; CDDP, cisplatin; GEM, gemcitabine; OS, overall survival; CBDCA, carboplatin; PTX, paclitaxel; VEGF, vascular endothelial growth factor; AB, antibody; PFS, progression-free survival; VNR, vinorelbine; AS, antisense oligonucleotide; FOLFIRI, irinotecan + 5-fluorouracil + leucovorin; XELOX, capcitabine + oxaliplatin; OX-CT, oxaliplatin-based chemotherapy; Iri-CT, irinotecan-based chemotherapy; Bev, bevacizumab; HER-2, human epidermal growth factor receptor-2; TTP, time to progression; RR, response rate; mTOR, mammalian target of rapamycin.

*Doxorubicin or epirubicin plus cyclophosphamide for patients who had never received an anthracycline or PTX for patients who had received adjuvant anthracycline.
therapies increased RR. Additionally, increase of RR appeared to be closely related to survival benefit in phase III trials. These results indicated that some level of objective response increment may be required, even in combination with standard therapies, for a successful drug in phase III studies of targeted agents.

In addition, the ROC analysis showed that the gain of RR was an accurate marker predicting, on average, survival benefit in past phase III trials with high sensitivity and specificity. These results may support the view that RR remains to be validated at least as an end point of randomized phase II studies for targeted agents in combination with standard therapies.

The regression model indicated that \( \Delta RR \) was required to be larger than the STE level of 21% or 15% if it is to be used in place of survival in phase III studies. We think this cut-off point is appropriate in the statistical view, but it may be too high when we discuss practical use of RR in screening studies. In screening studies, we think adequate sensitivity as well as high specificity (statistical accuracy) should be considered to avoid inappropriate rejection of agents that actually have meaningful clinical benefit. Therefore, we applied the ROC model in our analysis as a supporting data of the regression model. Because the binormal ROC model employed the maximum likelihood estimation to obtain the binormal parameters and optimized to reduce the difference between an actual population and the sample data [36], we believe that the ROC model was less effected by the size of cases in terms of the prediction bands or 95% confidence intervals. Considering the results of the ROC model, \( \sim 7\% \) might be appropriate for phase II studies to predict, on average, survival benefit in subsequent phase III studies. However, it should again be noted that for an individual trial to use an RR and impute, rather than measure, survival requires stronger datasets.

Some investigators have indicated that tumor response may not be a good surrogate for survival [37], and time-to-event end points such as PFS may be superior to RR in regard to surrogacy for OS [38]. As the recommendation of the use of randomized phase II design increased in oncology, especially
for the development of drug combinations [39], there may have been a corresponding increase in the use of PFS as a primary end point of phase II studies. In our analysis, HR of PFS appeared to be closely related to that of OS, and we do agree that PFS may be an appropriate predictor for OS, at least in phase III trials. However, small randomized trials bring with them unavoidable issues such as high false-positive and false-negative rates and unstable P values [40]. In particular, PFS has often been used as a primary end point in recent phase III studies, and therefore, the use of PFS as a single end point for a randomized phase II trial may render this an underpowered phase III trial. We believe that the use of RR, which is standardized and easily applicable to multicenter trials, in addition to PFS as an end point for randomized phase II studies could reduce these problems and give us further information which leads to successful phase III trials of MTA in combination with standard therapies. However, this interpretation requires a caution. We do not mean that RR can be used as a surrogate for survival in phase III trials. It requires a trial to yield ΔRR above the STE level of 21% or 15%, which may be a difficult test for a new drug. RR should be used as one of the screening measures in phase II studies, and thereafter, the presence of survival benefits should be investigated again in subsequent phase III studies. Our analysis has several limitations to be noted.

First, we have not covered a comprehensive review of all clinical trials including unpublished ones. Because negative studies would eventually stand a lower chance of being published, this publication bias can lead to an overestimation of treatment effect of targeted agents.

Second, the heterogeneity of study design in each trial may affect our results. The method of evaluating response is not exactly the same in each study (supplemental Table S1, available at Annals of Oncology online), and RR and PFS can be affected by these differences. Moreover, the method of sample size calculation is not the same across studies. Significance of survival benefit could be affected in part by the statistical property of such designs. In addition, there is a caution if we use the data of RRs in phase II studies. RRs observed in phase III studies may tend to be lower than those in phase II studies [41, 42].

Finally, there were four different cancer types included in our analysis. Biology and sensitivity to drugs may be different among each cancer type. These heterogeneities may influence the relation between response and survival benefit. However, there appeared to be no major differences among four cancer types in cut-off points of ARR predicting survival benefit. In addition, similar findings were observed in the linear regression analysis even if limited to lung cancer studies. Therefore, we believe our conclusion could be validated at least for lung studies.

In conclusion, our review indicated that RR increment obtained by the addition of MTA to a standard therapy was associated to survival benefit in clinical phase III trials of advanced cancer. The use of RR as an end point in screening studies of targeted agents in combination with standard therapies may help lead to successful phase III trials in terms of survival benefit, at least for lung cancer.

**funding**
The Osaka Medical Research Foundation for Incurable Diseases.

**acknowledgements**
We are grateful to Prof. Nagahiro Saijo (Department of Medical Oncology, Kinki University School of Medicine) for his critical
reading of the manuscript and his helpful comments. We thank Juri Mori for data management.

disclosure

None of the authors have any conflicts of interest associated with this study.

references


