VALUE IN HEALTH ▮ (2017) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

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

Do Surrogate End Points Better Correlate with Overall Survival in Studies That Did Not Allow for Crossover or Reported Balanced Postprogression Treatments? An Application in Advanced Non–Small Cell Lung Cancer

Mahmoud Hashim, MPH[1,*], *Boris M. Pfeiffer, PhD*[2], *Robert Bartsch, MSc*[1], *Maarten Postma, PhD*[3], *Bart Heeg, PhD*[1]

[1]*Ingress-health, Rotterdam, The Netherlands;* [2]*Merck KgaA, Darmstadt, Germany;* [3]*University of Groningen, Groningen, The Netherlands*

ABSTRACT

Background: In previous studies, correlation between overall survival (OS) and surrogate end points like objective response rate (ORR) or progression-free survival (PFS) in advanced Non-small cell lung cancer (NSCLC) was poor. This can be biased by crossover and postprogression treatments. **Objectives:** To evaluate the relationship between these two surrogate endpoints and OS in advanced NSCLC studies that did not allow for crossover or reported balanced post-progression treatments. **Methods:** A systematic review in patients with advanced NSCLC receiving second- and further-line therapy was performed. The relationship between the absolute difference in ORR or median PFS (mPFS) and the absolute difference in median OS (mOS) was assessed using the correlation coefficient (*R*) and weighted regression models. The analysis was repeated in predefined data cuts based on crossover and balance of postprogression treatments. When the upper limit of *R*'s 95% confidence interval (CI) was more than 0.7, the surrogate threshold effect (STE) was estimated. **Results:** In total, 146 randomized clinical trials (43,061 patients) were included. The mean ORR, mPFS, and mOS were 12.2% ± 11.2%, 3.2 ± 1.3 months, and 9.6 ± 4.1 months, respectively. The correlation coefficients of ORR and mPFS were 0.181 (95% CI 0.016–0.337) and 0.254 (95% CI 0.074–0.418), respectively, with mOS. Nevertheless, in trials that did not allow crossover and reported balanced postprogression treatments, the correlation coefficients of ORR and mPFS were 0.528 (95% CI 0.081–0.798) and 0.778 (95% CI 0.475–0.916), respectively, with mOS. On the basis of STE estimation, in trials showing significant treatment effect size of 41.0% or more ORR or 4.15 or more mPFS months, OS benefit can be expected with sufficient certainty. **Conclusions:** Crossover and postprogression treatments may bias the relationship between surrogate end points and OS. Presented STE calculation can be used to interpret treatment effect on either ORR or PFS when used as primary end points.

Keywords: crossover, non–small cell lung cancer, overall survival, surrogate end point validation.

Introduction

Overall survival (OS) is the criterion standard end point in cancer trials and is used to establish clinical benefit in support of regulatory and reimbursement applications [1–4]. Nevertheless, trials using OS as a primary end point need substantial sample sizes and extensive follow-up. In addition, the effects of crossover or unbalanced postprogression treatments may introduce bias or underestimate the treatment effect on OS [5,6]. An alternative surrogate end point for OS is progression-free survival (PFS). Regulatory agencies endorse PFS as a relevant end point in cancer trials [1,2,7]. In contrast to OS, PFS is not sensitive to postprogression treatments and has the advantage of assessing the duration of tumor response [5]. Objective response rate (ORR) is another potential surrogate end point. Compared with PFS, ORR does not assess response duration. The use of PFS and ORR as surrogate end points for OS would require that they be validated for this use [8]. Nevertheless, uncertainties regarding their association with OS and the potential for bias due to subjectivity in the assessment of ORR and PFS limit their use [7].

To our knowledge, only the Institute for Quality and Efficiency in Health Care (IQWiG) has issued a guidance document for surrogate end point validation in oncology [4]. The IQWiG recommends a stringent definition of surrogacy on the basis of the correlation coefficient (*R*). IQWiG states that if the lower limit of the 95% confidence interval (CI) of *R* is 0.85 or higher, validity of

* *Address correspondence to:* Mahmoud Hashim, Ingress-health, Health Economics and Real-World Evidence, Hofplein 20, Rotterdam 3032 AC, The Netherlands.

E-mail: mahmoud.hashim@ingress-health.com.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2017.07.011

the surrogate is suggested, but that the surrogate is not valid if the upper limit of the 95% CI is 0.7 or less [4]. Otherwise, the validity of the surrogate remains unclear; in this situation, IQWiG recommends estimating the surrogate threshold effect (STE) [4,9]. STE is defined as the minimum treatment effect on the surrogate necessary to predict a statistically significant nonzero effect on the true end point [9]. STE can be used to interpret the treatment effect on the surrogate end point.

A few studies in non–small cell lung cancer (NSCLC) have investigated the surrogacy of ORR or PFS to OS at the trial level [6,10–13]. These studies reported low correlations between PFS or ORR and OS. None of them included a stratified analysis based on the exclusion of studies allowing crossover or reporting unbalanced postprogression treatments. Stratifying studies on the basis of crossover has been done in other tumor types [14–16]. Delea et al. [15] assessed the surrogacy of PFS to OS in metastatic renal cell carcinoma trials. The correlation coefficient was greater in studies that did not allow/require crossover versus those that did allow/require crossover: correlation coefficients were estimated to be 0.50 and 0.28, respectively. Similarly, and to a less extent, greater correlation coefficients were observed in end point validation studies for metastatic melanoma and metastatic colorectal cancer after the removal of studies that did not allow/require crossover [14,16]. Hence, investigating the effect of crossover and postprogression treatments on the surrogacy of ORR or PFS to OS in NSCLC is warranted.

This study aimed to evaluate ORR and PFS as surrogate end points for OS in trials involving patients with advanced NSCLC receiving second- and further-line therapy. Then, the impact of crossover and unbalanced postprogression treatments on surrogacy was assessed.

Methods

Systematic Literature Review

The systematic literature review was conducted and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [17]. Two different bibliographic databases, PubMed and Embase, were used to identify published randomized clinical trials involving patients with stage IIIB/IV NSCLC receiving second- and further-line therapy. The search was conducted on July 28, 2016; no limitation on publication date was imposed.

A detailed search strategy (search syntax and eligibility criteria) is presented in Appendix Table 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.07.011. One investigator reviewed the titles/abstracts of retrieved articles sequentially using the predefined eligibility criteria (see Appendix Table 1 in Supplemental Materials). Subsequently, two investigators reviewed the full text of any article that appeared to meet the eligibility criteria; disagreement was resolved by consulting with a third investigator. References in publications reviewed at the full-text stage were evaluated to identify further relevant trials.

Upon agreement on the final list of included trials, one investigator extracted data from the included trials into a predefined Microsoft Excel template. Subsequently, another investigator validated the extracted data by re-extracting them. The following data were extracted: trial identification items (e.g., PubMed identifier, first author, year, trial phase, registration identifier, and trial acronym), interventions and target population, basic patient and disease characteristics (e.g., age, sex, performance status, disease stage, histology, metastasis, and number of previous lines of therapy), additional information (e.g., use of biomarkers and crossover), and data needed for end point validation (number of patients in each treatment arm, ORR, PFS, and OS). Risk of bias in individual studies was assessed using the Jadad scale [18].

Assessment of Publication Bias

The risk of bias across studies was assessed using funnel plots. In this study, trial size as a measure of precision was plotted on the y-axis, and treatment effect (absolute difference) on ORR, PFS, and OS was plotted on the x-axis. In the absence of publication bias, the plot should resemble a symmetrical inverted funnel [19].

Statistical Analysis

Primary analysis

The relationship between the absolute difference in ORR and median PFS (mPFS) and the absolute difference in median OS (mOS) was assessed using the correlation coefficient (R) and weighted linear regression models. A weighted linear regression model was fitted for the following two analyses: treatment effect on ORR, with the absolute difference in ORR (%) as an independent variable (predictor) and the treatment effect on OS (absolute difference in mOS in months) as a dependent variable; and treatment effect on PFS, with the absolute difference in mPFS (months) as an independent variable (predictor) and the treatment effect on OS (absolute difference in mOS in months) as a dependent variable. Analyses were weighted by trial size, as in previous end point validation studies [10,13,15,20–22].

Analyses were repeated using the absolute difference in ORR (%) or PFS hazard ratio (HR) and OS-HR because HRs might capture treatment effects not captured by median survival times. We carried out log transformation of HR. Log transformation can be used to make right-skewed distributions less skewed. Treatment effect on ORR is usually reported as the absolute difference in ORR (%). For that reason and for the ease of interpretation, we used it in both analyses with OS (mOS and OS-HR). Residual versus predicted plots were inspected and diagnostic tests for normality and heteroscedasticity (nonconstant error variance) were carried out to assess consistency with the assumptions of linear regression.

First, the analysis was conducted for all trials. Trials that had allowed crossover or in which postprogression treatments were unbalanced could underestimate OS benefit and subsequently bias surrogacy evaluation. Typically, phase III trials are adequately powered for end points such as PFS and OS, whereas phase II trials tend to be smaller and powered for safety end points or ORR. Thus, phase III trials might provide more information regarding the treatment effect on these end points. Therefore, second, on the basis of reported postprogression treatments, we examined trial-level surrogacy in all phase III trials (data cut A), in phase III trials excluding those with per-protocol crossover (data cut B), in phase III trials excluding those with both per-protocol and off-protocol crossover (data cut C), and in phase III trials excluding those with crossover, unbalanced postprogression treatments, or no information with regard to postprogression treatments (data cut D).

Trials that reported both the independent (the surrogate end point) and the dependent (the true end point) variables in both treatment arms were included in the analyses. For trials that included more than two treatment arms, the experimental arm was compared with a randomly chosen control arm within the same study to avoid analysis of correlated data, that is, including a treatment arm twice in the analysis. For trials that reported response in the evaluable population rather than in the intention-to-treat population, the denominator was adjusted to indicate the intention-to-treat population.

Assessing surrogacy and STE estimation

In cases in which the validity of the surrogate end point is deemed to be “unclear” following IQWiG guidelines [4], STE estimation is recommended to interpret treatment effect on the