VALUE IN HEALTH 17 (2014) 707–713

Available online at www.sciencedirect.com

ScienceDirect

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

Comparative Effectiveness Research/Health Technology Assessment (HTA)

Analyzing Overall Survival in Randomized Controlled Trials with Crossover and Implications for Economic Evaluation

Linus Jönsson, MD, PhD[1,*], *Rickard Sandin, PhD*[2], *Mattias Ekman, PhD*[1], *Joakim Ramsberg, PhD*[1], *Claudie Charbonneau, PhD*[3], *Xin Huang, PhD*[4], *Bengt Jönsson, PhD*[5], *Milton C. Weinstein, PhD*[6,7], *Michael Drummond, PhD*[8]

[1]*OptumInsight AB, Klarabergsviadukten, Stockholm, Sweden;* [2]*Global Health Economics and Outcomes Research, Pfizer Oncology, Sollentuna, Sweden;* [3]*Global Outcomes Research, Specialty Care BU, Pfizer, Paris, France;* [4]*Oncology Clinical Development, San Diego, CA, USA;* [5]*Stockholm School of Economics, Stockholm, Sweden;* [6]*Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA;* [7]*OptumInsight, Medford, MA, USA;* [8]*Centre for Health Economics, University of York, York, North Yorkshire, UK*

ABSTRACT

Background: Offering patients in oncology trials the opportunity to cross over to active treatment at disease progression is a common strategy to address ethical issues associated with placebo controls but may lead to statistical challenges in the analysis of overall survival and cost-effectiveness because crossover leads to information loss and dilution of comparative clinical efficacy. **Objectives:** We provide an overview of how to address crossover, implications for risk-effect estimates of survival (hazard ratios) and cost-effectiveness, and how this influences decisions of reimbursement agencies. Two case studies using data from two phase III sunitinib oncology trials are used as illustration. **Methods:** We reviewed the literature on statistical methods for adjusting for crossover and recent health technology assessment decisions in oncology. **Results:** We show that for a trial with a high proportion of crossover from the control arm to the investigational arm, the choice of the statistical method greatly affects treatment-effect estimates and cost-effectiveness because the range of relative mortality risk for active treatment versus control is broad. With relatively frequent crossover, one should consider either the inverse probability of censoring weighting or the rank-preserving structural failure time model to minimize potential bias, with choice dependent on crossover characteristics, trial size, and available data. A large proportion of crossover favors the rank-preserving structural failure time model, while large sample size and abundant information about confounding factors favors the inverse probability of censoring weighting model. When crossover is very infrequent, methods yield similar results. **Conclusions:** Failure to correct for crossover may lead to suboptimal decisions by pricing and reimbursement authorities, thereby limiting an effective drug's potential.

Keywords: cost effectiveness, crossover, oncology, sunitinib, surviva.

Introduction

Allowing patients the opportunity to switch to investigational therapy after the primary end point has been reached is a common strategy used to address ethical issues with the use of placebo-controlled randomized trials. This is common practice in oncology trials, often mandated by investigators, patients, and ethics committees [1]. Crossover can also occur when a trial is prematurely unblinded; for example, when an interim analysis shows a significant gain in the primary end point of the investigational treatment or if an active treatment (control or investigational) is less safe than its comparator. In each case, crossover results in loss of information about what the clinical effect would have been in the absence of crossover. A direct consequence of crossover is that standard statistical methods, for example, the intent-to-treat (ITT) analysis, may provide biased estimates of key end points such as overall survival (OS), for instance, and underestimation of the true effect. Furthermore, per protocol

Prior presentation: Jönsson L, Sandin R, Ekman M, et al. Analyzing overall survival in randomized controlled trials with cross-over. Presented at the ISPOR 13th Annual European Congress, Prague, Czech Republic, November 6–9, 2010.

Conflicts of interest: Linus Jönsson, Mattias Ekman, and Joakim Ramsberg were employees of OptumInsight (Stockholm, Sweden), who were paid consultants to Pfizer, Inc., in the development of this manuscript and production of the analyses. Joakim Ramsberg, Milton C. Weinstein, and Michael Drummond have all had an advisory role with Pfizer, Inc. Joakim Ramsberg has received honoraria from Pfizer, Inc., and Michael Drummond has received other remuneration from Pfizer, Inc. Rickard Sandin, Claudie Charbonneau, and Xin Huang are Pfizer, Inc., employees with stock ownership. Bengt Jönsson has no conflicts of interest to disclose.

* *Address correspondence to:* Linus Jönsson, Karolinska Institutet, SE-171 77 Stockholm, Sweden.

E-mail: linus.jonsson@ki.se.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2014.06.006

analysis (excluding patients who cross over to investigational treatment) can be subject to selection bias because those who switch from placebo to investigational treatment may not be representative of the entire placebo group. Design solutions such as randomizing crossover can be considered to minimize the impact of crossover [2] but will seldom be feasible and have rarely been implemented in practice. A more common approach is to adjust for crossover effects in the statistical analysis.

Crossover is also of concern in health economics and outcomes research because of its potential to affect estimates of efficacy and cost-effectiveness (CE). If an investigational drug reduces mortality, ITT analysis will underestimate the treatment effect in the presence of crossover and will likely lead to an overestimate of the incremental cost-effectiveness ratio (ICER). As a result, the decisions by pricing and reimbursement agencies regarding access to new therapies may not maximize health outcomes with the available resources if crossover is not corrected for.

In this article, directed toward policymakers, health care providers, health technology assessment agencies, and the pharmaceutical industry, we review available standard and advanced statistical methods for analyzing OS data in the presence of crossover and discuss choice of methodology. We illustrate differences between four methods with two case studies based on clinical trials in two indications for sunitinib (Sutent; Pfizer, Inc., New York, NY), an orally administered, multitargeted tyrosine kinase inhibitor, which is approved in several countries for the treatment of metastatic renal cell carcinoma (mRCC) and imatinib-resistant gastrointestinal stromal tumor (GIST) [3–8]. Both trials showed statistically significant benefit in progression-free survival (PFS) for sunitinib, and each illustrates a different situation regarding crossover. In the mRCC trial, crossover was infrequent and allowed only after overwhelming PFS results were observed [7], whereas in the GIST trial, crossover was frequent and allowed because of the placebo control design [4,9]. As we shall see, the two cases lead to interesting and potentially instructive differences in the use and outcomes of the four methods with respect to estimated treatment benefits and cost-effectiveness.

Statistical Analysis of Trials with Crossover

Statistical methods that are used to evaluate OS can be grouped into simple methods, which make no specific attempts to address crossover, and advanced methods based on statistical modeling techniques, which attempt to eliminate or reduce bias due to crossover.

Simple Methods

ITT analysis

In the standard ITT analysis, data for all randomized subjects are included for the entire period of observation. This method is appropriate as long as the aim is to compare one planned treatment with another irrespective of any subsequent treatment changes. If an investigational drug has a true mortality benefit, however, the ITT analysis will underestimate OS in the presence of crossover, and a cost-effectiveness analysis based on these data will likely overestimate the ICER of the new therapy [10].

Censoring at crossover (on-treatment analysis)

Censoring patients at crossover eliminates observations of patients randomized to the control arm after they receive the investigational treatment. Two disadvantages of the censoring method are selection bias and loss of power. Unless the probability of crossover is random, censoring may introduce bias, because events that result in censoring (e.g., progressive disease) may likely be associated with the final outcome (e.g., death) [2]. Censoring therefore leads to underestimation of gains in OS with the investigational treatment and selective exclusion of patients with a high probability of death. Censoring also lowers the power of the study because of shorter overall observation time and a reduced number of observed events in the control arm.

The potential selection bias induced by censoring can be reduced or eliminated if crossover is determined by randomization or if the entire control group crosses over to the investigational treatment at a prespecified point in time [2]. Even in these situations, however, censoring still reduces the statistical power of the study.

Statistical Modeling

During the last two decades, statistical modeling techniques have been developed to adjust for weaknesses in observational and clinical trial data. There is no criterion standard because the methods have different strengths and weaknesses. Two methods that have recently been used to attempt to correct for crossover in oncology trials are the inverse probability of censoring weighting (IPCW) model and the rank-preserving structural failure time (RPSFT) model. These apply statistical modeling techniques to reconstruct data for the control arm as if crossover had not occurred, with the aim of reducing bias and allowing the treatment effect to be assessed more accurately. Results based on these methods have been considered as relevant by health technology assessment bodies such as the National Institute for Health and Care Excellence (NICE) in the United Kingdom [11] and the Dental and Pharmaceutical Benefits Agency (tandvårds-och läkemedelsförmånsverket) in Sweden [12].

The IPCW model

The IPCW model is frequently used in epidemiologic research to adjust for nonrepresentative sampling or dropouts. In this method, patients who cross over from control to investigational treatment are censored, while patients remaining in the control arm are weighted to compensate for missing data [13]. The bias introduced by this informative crossover is corrected by weighting each patient by the inverse of his or her predicted probability of *not being censored* at a given time. The first step in the IPCW analysis is to predict the probability of crossover on the basis of each patient's baseline characteristics, such as age, sex, race, or biological markers [13–15], often by fitting a logistic regression model. Finally, OS is analyzed with the censored data set and observations weighted by the inverse of the predicted probability of censoring.

The IPCW model assumes that the probability of crossover at a given time depends only on observed covariates and must be independent of the outcome and its timing [13]. If these assumptions hold, then censoring can be made noninformative through the IPCW model. The clinical trial data must contain enough information about the covariates that affect the probability of crossover.

The RPSFT model

The RPSFT model allows a direct comparison of randomization groups by adjusting the OS of patients who cross over so that it reflects the OS had they not received the investigational treatment. The method is related to the accelerated failure time model in OS analysis [16,17], in which prognostic variables measured on the individual level are assumed to act multiplicatively on the time scale, for example, affecting the rate of progression. The first step is to define a causal model relating the *observed* event time T to the *unobserved* event time U that would have been observed if crossover had not occurred. This is performed by assuming that T