TREATMENT SWITCHING: STATISTICAL AND DECISION-MAKING CHALLENGES AND APPROACHES

Nicholas R. Latimer
School of Health and Related Research (ScHARR), University of Sheffield
n.latimer@sheffield.ac.uk

Chris Henshall
Brunel University London

Uwe Siebert
UMIT - University for Health Sciences, Medical Informatics and Technology Oncotyrol – Center for Personalized Cancer Medicine Harvard T.H. Chan School of Public Health and Massachusetts General Hospital, Harvard Medical School

Helen Bell
School of Health and Related Research (ScHARR), University of Sheffield

Objectives: Treatment switching refers to the situation in a randomized controlled trial where patients switch from their randomly assigned treatment onto an alternative. Often, switching is from the control group onto the experimental treatment. In this instance, a standard intention-to-treat analysis does not identify the true comparative effectiveness of the treatments under investigation. We aim to describe statistical methods for adjusting for treatment switching in a comprehensible way for nonstatisticians, and to summarize views on these methods expressed by stakeholders at the 2014 Adelaide International Workshop on Treatment Switching in Clinical Trials.
Methods: We describe three statistical methods used to adjust for treatment switching: marginal structural models, two-stage adjustment, and rank preserving structural failure time models. We draw upon discussion heard at the Adelaide International Workshop to explore the views of stakeholders on the acceptability of these methods.
Results: Stakeholders noted that adjustment methods are based on assumptions, the validity of which may often be questionable. There was disagreement on the acceptability of adjustment methods, but consensus that when these are used, they should be justified rigorously. The utility of adjustment methods depends upon the decision being made and the processes used by the decision-maker.
Conclusions: Treatment switching makes estimating the true comparative effect of a new treatment challenging. However, many decision-makers have reservations with adjustment methods. These, and how they affect the utility of adjustment methods, require further exploration. Further technical work is required to develop adjustment methods to meet real world needs, to enhance their acceptability to decision-makers.

Key words: Decision making, Statistical models, Survival analysis, Randomized controlled trials, Treatment switching, Treatment crossover

Treatment switching in a randomized controlled trial (RCT) occurs when patients in one group switch from the treatment specified for that group to the treatment specified for another group in the trial (most commonly from the control treatment to the experimental treatment) or to another treatment that is not part of the original trial protocol. The situation primarily arises due to ethical concern for the patient; the Declaration of Helsinki dictates that the goal of knowledge generation can never take precedence over the interests of individual research subjects (1). Therefore, if early analysis, or external information, suggests that clinical equipoise between the treatments under investigation is lost, there may be an ethical requirement to permit switching. The presence of switching in clinical trials creates difficulties in estimating the true effectiveness and cost-effectiveness of experimental treatments (2–4). This has implications for drug manufacturers, regulators, payers, clinicians,

and future patients: ultimately, a reliance on standard statistical techniques to analyze RCTs affected by treatment switching could result in effectiveness and harms being underestimated and access being denied to effective and cost-effective treatments. The policy issues around treatment switching are discussed in more depth in the accompanying study (5).

Statistical methods have been developed to adjust for the confounding introduced by treatment switching. The purpose of this study is to describe key adjustment methods and their main assumptions in a way that is understandable to nonstatisticians, and to summarize and discuss the implications of the views on the acceptability of these methods that were expressed by stakeholders at the 2014 Adelaide International Workshop on Treatment Switching in Clinical Trials.

BACKGROUND

The Adelaide workshop provided a platform for the discussion of the contentious issue of treatment switching. Stakeholder representatives from around the world were brought together with a view to identify where there is agreement on appropriate

approaches to address treatment switching in the design and analysis of oncology trials, and where further development is needed. The key stakeholders were identified as patients (and their representatives), oncologists, ethicists, statisticians, industry, regulators, health technology assessment (HTA) agencies, and payers. The main issues addressed at the Workshop included the design and management of oncology RCTs, the analysis, interpretation and validation of trial results, and their use in decision making.

The statistical adjustment methods can appear to be a complicated "black-box" to nonstatisticians. The fact that the workings of these methods are not widely understood has inhibited their acceptability, and therefore, usefulness. It became apparent during the Workshop that the general level of understanding of the adjustment methods needs to be improved, particularly in the HTA and payer communities. This study seeks to address this issue by explaining the principles of statistical adjustment methods and how they work in a way that is accessible to nonstatisticians. Diagrams are included to provide illustrated examples, but are presented only in the Supplementary Material for this study. We urge readers to refer to these diagrams to follow the explanations given for the adjustment methods; the diagrams are an integral part of the study.

ADJUSTING FOR TREATMENT SWITCHING

The Methods

A variety of adjustment methods were discussed at the Workshop, including their assumptions, strengths and weaknesses. This section provides an explanation of the approaches considered to be the most relevant for the switching observed in RCTs (2). We do not seek to explain all of the complexities of the methods; references are given for more detailed discussions available in the literature.

A standard intention-to-treat (ITT) analysis is generally accepted as the preferred approach for analyzing RCTs (6). The ITT analysis compares treatment groups as initially randomized, maintaining the balance generated by randomization and avoiding biases that can arise in intervention research such as nonrandom attrition of participants from the study (7;8). However, in the presence of treatment switching, the ITT analysis is not suited to the task of estimating the magnitude of the relative effect of the new treatment. If (as is usually the case) patients switch because they or their clinician believe that the experimental treatment has demonstrated superiority, the ITT analysis would generally be expected to provide an underestimate of the true effect of the experimental treatment compared with the control.

One approach to addressing the switching problem is to simply exclude patients who switch from the analysis, or to "censor" them from analysis at the point of switch. However, this approach is highly prone to selection bias, because switch-ers are likely to be prognostically different from nonswitchers (9–12). There is a general view that this "naïve" approach should be avoided (2–4).

Various more sophisticated adjustment methods have been developed to address the challenge of switching. Those currently considered most valid are (2,11–13): (i) marginal structural models (MSM) using inverse probability of censoring weighting (IPCW); (ii) two-stage adjustment; and (iii) rank preserving structural failure time models (RPSFTM) using g-estimation. These three approaches are described and discussed below.

Marginal Structural Models and Inverse Probability of Censoring Weights

It is simplest to think of the MSM with IPCW method as an extension of the simple censoring approach, whereby the bias associated with censoring is removed through the use of a weighting mechanism and the creation of a "pseudo population" (a population adjusted for the distortions that arise from the prognostic differences between switchers and nonswitchers) (14). The MSM with IPCW approach involves three key steps.

(1) First, switchers are censored at the point of switch. In Supplementary Figure 1a, we illustrate survival times observed for ten patients (five control group patients, and five experimental group patients) from a fictitious RCT, in which switching from the control group onto the experimental treatment was permitted after disease progression. In this example, three of the five control group patients switched, and are censored at the point of switch.

(2) Second, baseline and time-dependent information on patient characteristics that are prognostic factors for mortality and that influence the probability of switching are identified. These might include factors such as age and/or tumor burden and should be prespecified, and/or based on evidence or expert opinion. A dataset is then created such that each patient has a full set of values for these prognostic characteristics from randomization to the time of last contact. Models are fitted to this dataset to predict the probability of *not* switching in control group patients, conditional on the identified prognostic factors (Supplementary Figure 1b). These probabilities are used to determine the size of the weight applied to each patient.

Control group patients who did not switch but who have similar patient characteristics to patients who did switch are assigned higher weights, allowing them to account for themselves and for similar patients who switched. The weights may change for each patient over time, as the characteristics displayed by patients change. In our example, Patient 2, a control group patient who did not switch treatments, must have displayed dissimilar prognostic characteristics to patients who switched, because he or she received weights of less than 1.0 throughout his or her survival experience. In contrast, Patient 5 must have displayed similar prognostic characteristics to switchers, because he or she received a weight of greater than 1.0 after disease progression. All experimental group patients receive a weight of 1.0 throughout the MSM/IPCW analysis, because we do not wish to adjust for any switching in the experimental group (in this example). This step creates the (weighted) pseudo population.

(3) Finally, an analysis to estimate the adjusted treatment effect is undertaken (typically, a standard Cox regression model [15] is used to estimate a hazard ratio) (Supplementary Figure 1c). The Kaplan-Meier graph of survival probability against time presented in Supplementary Figure 1c presents the observed control group and experimental group

Kaplan-Meier curves, and the Kaplan-Meier curve of the weighted control group (i.e., the pseudo population).

The main challenge in using the MSM/IPCW approach is to ensure that the weightings compensate appropriately for the bias created by censoring switchers. For this to be fully achieved, data must be available at baseline and over time on all prognostic factors for mortality that also influence switching; this is known as the assumption of "no unmeasured confounding" (14–16). We can never be sure that this assumption holds, because we do not have perfect knowledge on which characteristics are prognostic. However, a clear statement of which prognostic characteristics have been considered and why (e.g., based upon data from other trials and/or clinical judgement), and comprehensive data collection on these may help to re-assure decision makers that the result of an MSM/IPCW analysis is credible.

A further problem is that, although it may be desirable to include as many potentially prognostic covariates as possible within the MSM/IPCW analysis, in relatively small RCTs this may lead to convergence issues, that is, the method does not produce a result. Also, specific patients may be assigned very high weights, particularly if there are very few control group patients who did not switch treatments. In this case, the method becomes prone to substantial error, because there is a higher likelihood that these remaining patients may not be representative of the wider trial population.

Two-Stage Adjustment

The Two-Stage adjustment method to address the treatment switching problem was developed for the situation that often pertains in RCTs of oncology treatments, where switching only occurs after disease progression (2;12;13). The method involves three steps, which allow survival times that would have been observed in the absence of switching to be estimated (these are also known as "counterfactual" survival times).

(1) First, a "secondary baseline" is identified for patients in the control group (assuming that we wish to adjust for switching only in the control group). Switching must only occur after this secondary baseline. Supplementary Figure 2 provides an illustration of a Two-Stage adjustment analysis applied to the fictitious RCT previously introduced. In this example, the point of disease progression is used as the secondary baseline for each patient (Supplementary Figure 2a).

(2) Second, postsecondary baseline (in this case post-disease progression) survival is analyzed within the control group (Supplementary Figure 2b). A treatment effect (in the form of an "time ratio") associated with the experimental treatment is then estimated specifically for control group switching patients, compared with control group nonswitchers, using a model that is adjusted for patient characteristics measured at the time of the secondary baseline. Patients are assumed to be at a similar stage of disease at the secondary baseline, and if prognostic characteristics are measured at this time-point we can adjust for any differences between switchers and nonswitchers, and thus produce an unbiased estimate of the treatment effect in switchers. The prognostic characteristics considered here are similar to those considered in the MSM/IPCW method previously described, but for the Two-Stage method, data are only required on these at the time of disease progression, not beyond.

(3) Third, the treatment effect associated with switching is used to "shrink" the survival times observed in switchers, to provide estimates of the survival times that would have been observed in the absence of switching (Supplementary Figure 2c). The adjusted control group survival times are then compared with observed experimental group survival times to obtain an estimate of the adjusted treatment effect.

This approach is reliant on switching occurring after a specific disease-related time-point, so that a secondary baseline can be defined. At this time-point, the assumption of "no unmeasured confounders" must hold (information must be available on all patient characteristics that are associated with survival), creating similar challenges to those discussed for this assumption in the previous section. A further limitation of the Two-Stage method is that it only adjusts for differences between switchers and nonswitchers at the time of the secondary baseline: it does not capture any changes in patient characteristics that may occur between the secondary baseline and the time of switch. Hence, switching must occur soon after the secondary baseline, otherwise differences between switchers and nonswitchers may not be adequately adjusted for.

The Two-Stage adjustment method is prone to convergence issues if trial sample sizes are small, particularly if there are several important prognostic characteristics. Also, estimates of the treatment effect in switchers will be prone to error if there are very few control group patients who do not switch. However, research has suggested that the Two-Stage adjustment method is less sensitive to these issues than the MSM/IPCW (12;13).

A further potential issue with the Two-Stage method arises around something called "re-censoring." It is not discussed here, due to its complexity, but it is discussed at length elsewhere (2;17), and is summarized in the Supplementary Material for this study. Importantly, the Two-Stage method can be applied with or without re-censoring and either approach could be prone to bias, depending upon the characteristics of the trial and the treatment in question.

Rank Preserving Structural Failure Time Models (RPSFTM)

Like the Two-Stage method, the RPSFTM estimates survival times that would have been observed had treatment switching not occurred (18). The method involves three steps.

(1) First, for each patient, survival is split into two segments: time spent "on" the experimental treatment (referred to as T_{On_i}), and time spent "off" the experimental treatment (referred to as T_{Off_i}), defined according to treatment initiation and discontinuation times for each patient. In Supplementary Figure 3a, this is illustrated for the fictitious RCT previously introduced.

(2) The second step is reliant upon two key assumptions. It is assumed that if no patients in either trial group had received the experimental treatment, the average survival time in the two groups would have been equal because the two groups were created through randomization. It is further

assumed that all patients, whether in the control or experimental group, receive the same degree of benefit (treatment effect, or what is referred to as $e^{-\psi_0}$ in the formula presented in Supplementary Figure 3b) from their time on experimental treatment, known as the "common treatment effect" assumption. Survival is, therefore, a simple function of time on and off the treatment and the treatment effect (Supplementary Figure 3b).

Given these assumptions, a process called g-estimation is used to identify the treatment effect. G-estimation essentially amounts to a grid search of possible values for the treatment effect. Using the formula presented in Supplementary Figure 3b, for each patient the inverse of the treatment effect (i.e., e^{ψ_0}) is applied as a factor to the time spent on treatment, to nullify the effect of having received the treatment, hence providing untreated survival times (referred to as U_i in Supplementary Figure 3b). Once this has been done for all patients, untreated survival times are compared between randomized groups; the "true" value of the treatment effect is the one that results in equal untreated survival times between the randomized groups.

(3) Once the "true" value of the treatment effect has been identified, adjusted survival times can be calculated for the control group and compared with the observed experimental group survival times to obtain an estimate of the treatment effect adjusted for treatment switching (Supplementary Figure 3c).

The assumption that randomization has "worked" should not be problematic in an RCT, and is also required by standard ITT analyses. The common treatment effect assumption may be more problematic, if the capacity to benefit from a treatment is dictated by when it is received. In oncology trials, treatment switching often occurs after some defined point of disease progression, and it may not be plausible to assume that switchers will attain the same benefit from treatment (relative to the time it is taken for) as those who received the treatment immediately upon randomization. More complex versions of the RPSFTM, which allow the common treatment effect assumption to be relaxed, have proven unsuccessful (17;19;20).

The structure of the RPSFTM model suggests that patients must be either "on" or "off" treatment, if the control treatment is active this may be problematic because it may be possible to be on treatment x, on treatment y, or off treatment. The RPSFTM can still be applied in these circumstances, but requires additional assumptions (2).

Simulations have shown that the RPSFTM performs extremely well when the common treatment effect assumption holds (11–13). Also, RPSFTMs are less sensitive to issues arising from small patient numbers than the Two-Stage adjustment and IPCW methods (12;13). However, problems can occur and the g-estimation procedure may not work well with very small patient and event numbers. Like the Two-Stage method described in the previous section, the RPSFTM can be applied with or without re-censoring (see the Supplementary Materials for information on re-censoring).

THE VIEW OF DECISION MAKERS

While it has been reported that statistical approaches to adjust for treatment switching in RCTs can provide reliable estimates of causal treatment effects in a range of scenarios (2;11–13), it was clear from discussions at the Workshop that the confidence placed in these adjustment analyses differs substantially between stakeholders. It appears that regulators are unlikely to consider adjustment analyses sufficiently convincing on their own for licensing purposes except in particular situations such as when trials have a low degree of switching and/or where there is substantial supportive evidence from other sources. While some payers have used adjustment analyses to inform recent decisions, many have not and all appear to have reservations. Some take these as grounds for ignoring adjustment analyses entirely, either considering only ITT analyses (despite their known inadequacies) or ignoring the trials entirely. Others attempt to factor adjustment analyses into their wider appraisal of evidence. Issues around the use of adjustment analyses appear to fall into two broad categories: concerns over methodological assumptions and confidence intervals, and issues around how best to use adjustment analyses alongside other relevant information in the decision-making process.

Concerns over Methodological Assumptions and Confidence Intervals

Each of the adjustment methods makes important assumptions. While some of these may appear to lack face-validity in some cases, analyses can be undertaken to provide information on their validity for specific situations. Stakeholders at the Workshop agreed that if adjustment methods were used, the following types of analyses should be undertaken to establish the rationale for each method on a case-by-case basis.

Investigation of the "No Unmeasured Confounders" Assumption. The assumption of "no unmeasured confounders" is pivotal for the MSM/IPCW and Two-Stage adjustment methods. It is impossible to test this assumption fully, but information can be obtained from other trials in similar disease areas to determine whether any likely prognostic characteristics were not measured in the pivotal trial. Clinical expert opinion could also be sought on its plausibility (2). Discussion at the Workshop suggested that prognostic covariates (both for survival *and* for switching) should be identified as far as possible during trial planning, and data should be collected on these during the trial. Importantly, this data collection should be continued beyond disease progression to enhance the likelihood that the full range of adjustment methods can subsequently be applied.

Investigation of the Likely Bias Associated with Two-Stage Adjustment. If there is a significant time-lag between the secondary baseline and the treatment switch, the Two-Stage adjustment method becomes prone to bias (2). Hence, analyses of time-to-switch should be undertaken, to provide information on the scope for bias. For instance, if an analysis of the time from secondary baseline to time of switch demonstrates that all switchers began to receive the experimental treatment very soon after disease progression (in a matter of days), the scope for time-dependent

confounding is minimal. In contrast, if patients switched months or years after the secondary baseline the scope for time-dependent confounding is much higher.

Investigation of the "Common Treatment Effect" Assumption. The "common treatment effect" assumption is pivotal for RPSFTM. It is impossible to test this assumption fully, but attempts should be made to assess its validity. This could be achieved, for instance, by using the Two-Stage adjustment method to derive and compare treatment effect estimates for switchers and for patients randomized to the experimental group. Also, clinical expert opinion could be sought to determine the biological plausibility of a common treatment effect (2). If the common treatment effect assumption is deemed implausible, the extent to which the treatment effect differs between groups could be investigated to provide information on the potential size of the bias associated with a violation of this assumption.

Investigation of Other Factors. Several other factors should be investigated:

(1) Adjustment methods (particularly the MSM/IPCW) become prone to higher levels of error when the switching proportion is very high and/or the sample size in strata/cells defined by combinations of covariates used within the MSM is small and/or the number of events observed in these strata/cells is low (2;12;13;21). Details on each of these should be presented to allow reviewers to assess the scope for bias.

(2) It is possible to apply the RPSFTM under the assumption of a treatment effect that either stops or continues at the point of treatment discontinuation, and to apply the RPSFTM and Two-Stage adjustment methods with or without re-censoring (there are circumstances where re-censoring itself can lead to bias) (2;17). The appropriateness of each of these options should be investigated by studying the treatment effect over time, and the impact of re-censoring.

(3) The most common perceived need for adjustment is to address switching from the control group onto the experimental treatment. However, if other switches occur, in either group, consideration should be given to adjusting for these: if we adjust for switching in the control group, it may be argued that we should also adjust for switches from the experimental group onto other investigational or nonstandard treatments.

Confidence Intervals. In addition to the assumptions involved in an adjustment analysis, there may be concerns about the confidence intervals of the estimates they generate. For the RPSFTM, it should be recognized that often confidence intervals will widen, compared with the ITT analysis, primarily as a result of the method design: If the ITT analysis does not provide a statistically significant result, the equivalent RPSFTM analysis also will not, because typically the ITT analysis *p* value is retained. In contrast, an MSM/IPCW may result in narrower confidence intervals, though this is dependent upon the covariate data available. For regulators and HTA agencies, this uncertainty may be important. For those only willing to consider statistically significant results, the standard application of the RPSFTM may be of limited utility, particularly in circumstances where the ITT analysis does not produce a statistically significant result.

Using Adjustment Analyses and Other Relevant Information in Decisions

Although decision makers appear to be reluctant to place reliance on methods that they and/or their stakeholders do not fully understand, they appear to accept that ignoring the likely effects of switching in trials may lead to underestimation of the extent of benefits and possible harmful effects and, in some cases, decisions that deny patients access to effective products. There is, therefore, an interest in using adjustment analyses as an indicator of the likely extent of confounding that switching gives rise to, and to at least consider such analyses in an overall judgement of the likely comparative effectiveness and cost-effectiveness of a new treatment. Those making such judgements may also find it helpful to see data from sources outside the pivotal trial.

Sources of relevant data from outside the main trial include registry data (retrospective or prospective alongside the trial) or data from other trials that may inform estimates of long-term survival unaffected by switching (2;22). External data from relevant RCTs unaffected by treatment switching that included the same endpoints as the trial under investigation are likely to be particularly useful.

It was recognized in discussions that using external or observational data is likely to give rise to methodological issues, such as population and protocol differences and the age of the data. But such analyses may still represent useful supporting evidence, particularly if patient-level data can be obtained such that adjustments can be made for differences in patient characteristics. Again, it was clear from discussions at the Workshop that the acceptability of combining data from multiple sources varied between stakeholders, reflecting concerns regarding validity and relevance.

PROPOSALS FOR GOOD PRACTICE IN RELATION TO TREATMENT SWITCHING

Several strong messages around the analysis of trials affected by treatment switching emerged from the Workshop:

(1) In the presence of treatment switching that is deemed to materially affect the results of a trial, a standard ITT analysis may usefully be supplemented with analyses that attempt to adjust for switching and which may help to address the particular decision problems faced by different stakeholders.

(2) Where adjustment methods are used, trial characteristics and the treatment switching mechanism should be considered alongside the pivotal assumptions associated with each of the statistical adjustment methods to identify which of the analyses are likely to provide the least amount of bias.

(3) The sensitivity of results should be tested through application of all methods deemed to be potentially appropriate: If a range of methods are shown to be potentially appropriate for a particular case, and each provides similar estimates of the treatment effect, decision makers may have more confidence in the results.

(4) Appropriate use of external data may be explored, to provide further supportive evidence.

FURTHER RESEARCH

Important limitations with adjustment methods remain, and various reservations regarding their use for decision-making purposes are legitimate. Barriers to the use of adjustment methods largely relate to a mis-trust of methodological assumptions and an inability by some fully to understand the theoretical basis for the methods (and, therefore, to appraise their strengths and weaknesses). Further exploration of the specific reservations held by different decision makers may provide the catalyst for a provision of education to address some of these, while also highlighting areas where further methodological research is needed, conducted in the light of real world decision making contexts and constraints.

Discussions with international stakeholders in follow-up to the Adelaide meeting are ongoing, which will hopefully allow research areas to be identified in a timely manner. Areas of particular value are likely to address formal methods for assessing the validity of the key assumptions made by the adjustment methods, and methods for integrating evidence from external trials or observational studies to estimate what survival would have been in the absence of switching. Further research would permit more detailed guidance on how to identify an appropriate adjustment method for a specific trial, and how the design of trials and wider clinical development programs might better be adapted to support the use of adjustment methods.

CONCLUSIONS

There is a need for decision makers to openly engage with the use of adjustment methods in the presence of treatment switching in RCTs. A standard ITT analysis is highly likely to produce a biased estimate of the comparative treatment effect in the presence of switching (typically underestimating benefits and possible harms), and although adjustment methods are not perfect, evidence suggests that they will often provide an estimate of the treatment effect that is closer to the truth (2;11–13). It is important that trials provide information that addresses the needs of decision makers, otherwise their ethical justification is questionable. It is hoped that this study and its companion (5) will help trial sponsors plan, manage and analyze trials where treatment switching can be expected so as to provide better information for decision makers.

It can also be argued that there is an ethical obligation for decision makers to make appropriate use of information from trials for which patients have volunteered in the belief that they will be contributing to better decisions, taking due account of the possible shortcomings of such information. It is hoped that this study, and further research on this topic, will help decision makers understand the value and limitations of adjustment methods and how these can be used to analyze trials that are relevant for decisions but biased by treatment switching. It is further hoped that this will lead to increased appropriate use of these methods and to better informed decisions.

SUPPLEMENTARY MATERIAL

Supplementary Figure 1: http://dx.doi.org/10.1017/S026646231600026X

Supplementary Figure 2: http://dx.doi.org/10.1017/S026646231600026X

Supplementary Figure 3: http://dx.doi.org/10.1017/S026646231600026X

Supplementary Description of re-censoring: http://dx.doi.org/10.1017/S026646231600026X

REFERENCES

1. World Medical Association. Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. *JAMA*. 2013;310: 2191-2194.
2. Latimer NR, Abrams KR, Lambert PC, et al. Adjusting survival time estimates to account for treatment switching in randomised controlled trials – An economic evaluation context: Methods, limitations and recommendations. *Med Decis Making*. 2014. doi:10.1177/0272989X13520192.
3. Jonsson L, Sandin R, Ekman M, et al. Analyzing overall survival in randomized controlled trials with crossover and implications for economic evaluation. *Value Health*. 2014;17:707-713.
4. Ishak KJ, Proskorovsky I, Korytowsky B, et al. Methods for adjusting for bias due to crossover in oncology trials. *Pharmacoeconomics*. 2014;32:533-546.
5. Henshall C, Latimer NR, Sansom L, Ward R. Treatment switching in clinical trials: Issues and proposals following the 2014 Adelaide International Workshop, *IJTAHC* 2016;32:167-174.
6. Moher D, Schulz KF, Altman DG. CONSORT GROUP (Consolidated Standards of Reporting Trials). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med*. 2001;134:657-662.
7. Wertz RT. Intention to treat: Once randomized, always analyzed. *Clin Aphasiol*. 1995;23:57-64.
8. Gupta SK. Intention-to-treat concept: A review. *Perspect Clin Res*. 2011;2:109-112.

9. Lee Y, Ellenberg J, Hirtz D, Nelson K. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med.* 1991;10:1595-1605.

10. Horwitz R, Horwitz S. Adherence to treatment and health outcomes. *Arch Intern Med.* 1993;153:1863-1868.

11. Morden JP, Lambert PC, Latimer NR, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. *BMC Med Res Methodol.* 2011;11:4.

12. Latimer N, Abrams K, Lambert P, et al. Adjusting for treatment switching in randomised controlled trials – A simulation study and a simplified two-stage method. *Stat Methods Med Res.* 2014. doi:10.1177/0962280214557578

13. Latimer NR, Abrams KR, Lambert PC, Morden JP, Crowther MJ. Assessing methods for dealing with treatment switching in clinical trials: A follow-up simulation study. *Stat Methods Med Res.* 2016. doi:10.1177/0962280216642264

14. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Statist Assoc.* 2001;96:440-448.

15. Cox DR. Regression models and life tables (with Discussion). *J R Stat Soc Series B Stat Methodol.* 1972;34:187-220

16. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics.* 2000;56:779-788.

17. White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: Examples from the Concorde trial. *Stat Med.* 1999;18:2617-2634.

18. Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. *Commun Stat Theory Methods.* 1991;20:2609-2631.

19. Robins JM, Greenland S. Adjusting for differential rates of prophylaxis therapy for Pcp in high-dose versus low-dose Azt treatment arms in an aids randomized trial. *J Am Stat Assoc.* 1994;89:737-749.

20. Yamaguchi T, Ohashi Y. Adjusting for differential proportions of second-line treatment in cancer clinical trials. Part II: An application in a clinical trial of unresectable non-small-cell lung cancer. *Stat Med.* 2004;23:2005-2022.

21. Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *Am J Epidemiol.* 2011;173:569-577.

22. Drummond M, Evans B, LeLorier J, et al. Evidence and values: Requirements for public reimbursement of drugs for rare diseases – A case study in oncology. *Can J Clin Pharmacol.* 2009;16:e273-e281.