Accounting for Uncertainty in Decision Analytic Models Using Rank Preserving Structural Failure Time Modeling: Application to Parametric Survival Models

Iain Bennett, MSc[1],*, Noman Paracha, MSc[1], Keith Abrams, PhD (Professor)[2], Joshua Ray, MSc[1]

[1]F. Hoffmann-La Roche AG, Basel, Switzerland; [2]University of Leicester, Leicester, UK

A B S T R A C T

Objectives: Rank Preserving Structural Failure Time models are one of the most commonly used statistical methods to adjust for treatment switching in oncology clinical trials. The method is often applied in a decision analytic model without appropriately accounting for additional uncertainty when determining the allocation of health care resources. The aim of the study is to describe novel approaches to adequately account for uncertainty when using a Rank Preserving Structural Failure Time model in a decision analytic model. **Methods:** Using two examples, we tested and compared the performance of the novel Test-based method with the resampling bootstrap method and with the conventional approach of no adjustment. In the first example, we simulated life expectancy using a simple decision analytic model based on a hypothetical oncology trial with treatment switching. In the second example, we applied the adjustment method on published data when no individual patient data were available. **Results:** Mean estimates of overall and incremental life expectancy were similar across methods. However, the bootstrapped and test-based estimates consistently produced greater estimates of uncertainty compared with the estimate without any adjustment applied. Similar results were observed when using the test based approach on a published data showing that failing to adjust for uncertainty led to smaller confidence intervals. **Conclusions:** Both the bootstrapping and test-based approaches provide a solution to appropriately incorporate uncertainty, with the benefit that the latter can implemented by researchers in the absence of individual patient data.

Copyright © 2017, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Modern randomized controlled trials (RCTs), which remain the gold standard in terms of evaluating the efficacy and safety of new interventions, often accommodate treatment switching from the control group to the experimental treatment group at some point during the trial. Treatment switching is primarily driven by ethical considerations; for instance, it would be unethical to disallow treatment switching for patients randomly allocated to therapy shown to be inferior in an interim analysis, particularly in cases where no other nonpalliative therapy options exist. Moreover, treatment switching can be used to boost trial recruitment, for example, by allowing switching after a primary endpoint has been observed (commonly, progression-free survival) [1]. It has been reported that over half the recent health technology assessments (HTAs) in oncology performed by the National Institute for Health and Care Excellence (NICE) in England and Wales and the Pharmaceutical Benefits Advisory in Australia have involved trials that included treatment switching [2].

Standard statistical approaches used in the analysis of RCTs are designed to compare groups based on the intention-to-treat (ITT) principle, which means that patients are analyzed according to their randomized treatment assignment and that all patients who were enrolled and received treatment are included in the analysis [3]. When patients in both groups receive the investigational intervention in a trial, such conventional analyses may not provide an accurate estimate of the comparative effectiveness of the two therapies, particularly for endpoints, such as overall survival (OS), which is critical for cost-effectiveness analysis, even though it is often not the primary endpoint of the trial. Although it is ethically justifiable to allow patients to switch to an experimental therapy after reaching the primary endpoint (e.g., progression-free survival), which may be the key endpoint for regulatory approval, methods are required to adjust for the effects of treatment switching on other endpoints (e.g., OS) that are crucial for health economic analysis and HTA decision making.

Simple methods of adjusting for treatment switching have been historically used in HTAs, such as those excluding switchers from the analysis or censoring their data at the time of switch, but these can create selection bias because treatment switching is often related to prognosis [4]. Recent recommendations indicate that these simple approaches should be avoided for the

---

http://dx.doi.org/10.1016/j.jval.2017.07.008

estimation of OS and replaced with methodologies that preserve randomization and are designed to address the issue of bias instead [5]. The Rank Preserving Structural Failure Time (RPSFT) model, inverse probability of censoring weights and two-stage adjustment estimation methods have all been shown to produce unbiased adjustments, provided the assumptions underpinning them hold true [6–8]. The RPSFT method, introduced by Robins and Tsiatis, provides an estimate of the OS time for the control group had treatment switching not occurred [6]. It estimates OS measured from the time of treatment switching by applying an estimate of the benefit of the experimental treatment (derived iteratively and referred to as the *inverse of the acceleration factor*). This method assumes that the benefit of the experimental treatment is the same whether it was received from the time of randomization or only received later as a switch treatment (referred to as the "constant treatment effect" assumption).

Given the potential confounding caused by treatment switching, it is important that appropriate adjustment methods are used for health economic analyses based on treatment switching trials and for informing HTAs. For example, in a 2012 NICE appraisal of vemurafenib for the treatment of melanoma the incremental cost-effectiveness ratio was decreased from over £75,000 per quality-adjusted life year gained to approximately £52,000 after adjusting for treatment switching [9]. This evidence suggests that failure to appropriately adjust for treatment switching has the potential to lead to misinformed HTA decision making. Although the use of adjustment methods in HTA submissions is beginning to be accepted in some countries, there is paucity of data on the role of adjustment methods in probabilistic sensitivity analysis (PSA), which is used to capture uncertainty and inform decision making in the HTA process [2,10]. PSA can be defined, in terms of a health economic modeling analysis, as the process in which "all input parameters are considered as random quantities and therefore are associated with a probability distribution that describes the state of science" [11]. The most commonly used adjustment method (RPSFT model) is known to introduce additional uncertainty when estimating (adjusted) hazard ratios (HRs) in treatment switching trials, an effect which also has the potential to influence HTA decision making [6]. When survival times are adjusted for treatment switching within decision analytic models, these adjusted HRs are rarely used explicitly. Instead, more commonly parametric survival curves are generated based on the adjusted patient survival.

The aim of the present study was to describe novel approaches to adequately adjust for uncertainty when using an RPSFT model, by (1) simulating life expectancy using a straightforward decision analytic model based on a hypothetical oncology trial with treatment switching, and (2) applying one of the approaches on published data to demonstrate the value of adjusting for uncertainty when using RPSFT models.

Methods

In a standard application of RPSFT model proposed by Robins and Tsiatis [6], two different survival times for a patient, i, are considered with notation:

- T_i– the observed survival time
- U_i– the latent survival time with no treatment

An accelerated failure time model is proposed to relate these, such as:

$$U_i(\varphi) = T_{Ci} + T_{Ei}e^{\varphi}$$

where T_{Ei} is the observed time on experimental therapy, and T_{Ci} is defined as $T_{Ci} = T_i - T_{Ei}$. The treatment parameter theta (φ) is an unknown with true value φ_0. By assuming the latent survival times will be balanced through randomization a g-estimation procedure can be used to estimate φ_0. This g-estimation proceeds by proposing a candidate set of values for the unknown parameter φ; estimating the latent survival time $U_i(\varphi)$ for both arms and then comparing as randomized using a suitable test. The candidate value of φ which leads to no difference in the comparison of the latent survival time as randomized is then taken as the estimate for φ_0. Using this estimate for φ_0 the counter factual latent survival U_i for the control arm can then be compared with the observed survival time T_i of the experimental arm using standard statistical methods. Robins and Tsiatis noted that when considering confidence intervals (CIs) for HRs estimated from RPSFT corrected data, the P value from the test used in the g-estimation procedure should be used to create symmetric CIs [6]. This is typically done by estimating an adjusted standard error (SE_{ADJ}) for the treatment effect (β) using equation 1, where X_{ITT}^2 is the chi-square statistic from the log rank test used for g-estimation applied to the ITT comparison.

$$SE_{ADJ}(\hat{\beta}) = \frac{\hat{\beta}}{\sqrt{X_{ITT}^2}} \tag{1}$$

The present analysis describes an extension to this approach for use in parametric extrapolation and comparison with the alternative approach of bootstrapping with a small simulation study. The simulation study and a reanalysis of published data are used to illustrate the impact of not performing such a correction on PSA in a decision analytic model.

Estimating the Covariance Matrix Using an Adjustment Factor

The method assumes that an RPSFT model has been used to estimate counterfactual survival times for patients on standard care and assuming that treatment switching had not occurred following the approaches described in detail in the literature [6,8,12]. Following this step, the algorithm to apply the adjustment to a parametric covariance matrix is as follows:

1. Fit a parametric survival model to the observed data for the experimental arm and the counterfactual control arm survival, including a treatment effect parameter (coded to indicate being on control arm therapy relative to experimental arm, so the intercept represents the effect of experimental treatment). This is done so all the additional variance from the RPSFT method is contained in the treatment effect and is not split between the treatment effect and the intercept.

2. Derive an adjustment factor using equation 2, where $SE_{ADJ}(\beta)$ is defined as with equation 1, where β is the estimate of the treatment effect from the parametric model, and $SE_{OBS}(\beta)$ is the estimated standard error.

$$F = \frac{SE_{ADJ}(\beta)}{SE_{OBS}(\beta)} \tag{2}$$

3. Multiply all components of the covariance matrix that involve covariance with treatment effect by this adjustment factor. This assumes that the correlation between the parametric model parameters and treatment effect is not modified through the use of the RPSFT adjustment. This is illustrated for a Weibull model with parameters μ (intercept), β (treatment effect for control relative to experimental) and γ (shape) in equations 3 and 4; however, very similar derivations apply for other parametric survival models. Equation 3 shows the

estimated covariance matrix, whereas equation 4 shows the form of the adjusted matrix using the proposed method.

$$V = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = \begin{bmatrix} \sigma_\mu^2 & \rho_{(\mu\beta)}\sigma_\mu\sigma_\beta & \rho_{(\mu\gamma)}\sigma_\mu\sigma_\gamma \\ \rho_{(\mu\beta)}\sigma_\mu\sigma_\beta & \sigma_\beta^2 & \rho_{(\beta\gamma)}\sigma_\beta\sigma_\gamma \\ \rho_{(\mu\gamma)}\sigma_\mu\sigma_\gamma & \rho_{(\beta\gamma)}\sigma_\beta\sigma_\gamma & \sigma_\gamma^2 \end{bmatrix} \quad (3)$$

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} = \begin{bmatrix} \sigma_\mu^2 & \rho_{(\mu\beta)}\sigma_\mu \boldsymbol{F}\sigma_\beta & \rho_{(\mu\gamma)}\sigma_\mu\sigma_\gamma \\ \rho_{(\mu\beta)}\sigma_\mu \boldsymbol{F}\sigma_\beta & (\boldsymbol{F}\sigma_\beta)^2 & \rho_{(\beta\gamma)}\boldsymbol{F}\sigma_\beta\sigma_\gamma \\ \rho_{(\mu\gamma)}\sigma_\mu\sigma_\gamma & \rho_{(\beta\gamma)}\sigma_\beta\sigma_\gamma & \sigma_\gamma^2 \end{bmatrix}$$

$$= \begin{bmatrix} v_{11} & Fv_{12} & v_{13} \\ Fv_{21} & F^2 v_{22} & Fv_{23} \\ v_{31} & Fv_{32} & v_{33} \end{bmatrix} \quad (4)$$

This adjusted covariance matrix can then be used for PSA in the usual manner following standard methods as described by Briggs et al. [13].

Estimating the Covariance Matrix with Bootstrapping

Alternatively, to generate a covariance matrix with bootstrapping, samples from the data set are taken with replacement, and the entire RPSFT procedure, including grid search is repeated. The parametric model is then fit to the observed experimental data and counterfactual survival data to yield a set of parameters for each sampled data set. The covariance of these bootstrapped parameters is then used as an estimate for the covariance matrix (Fig. 1). It should be noted that bootstrapping of g-estimation procedures is not trivial because considerable care has to be taken in how the convergence of the bootstrap estimates is assessed and nonconvergence included in the results.

Simulation Study Design

To illustrate the performance of the novel test-based approach compared with bootstrapping the covariance matrices and to making no adjustment to the covariance matrix, 500 data sets simulating a common RCT design were generated. The simulation design used was identical to that of Scenario Number 1, which was described in detail by Morden et al. and which was believed to reflect what is often observed in clinical practice [14]. Each data set contained 500 patients randomly allocated in a 1:1
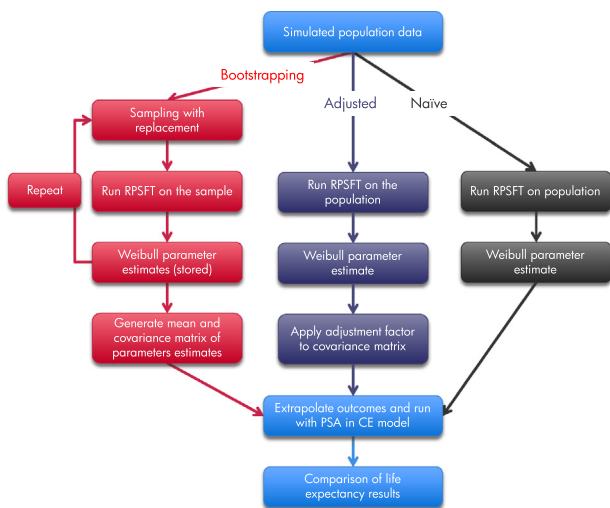

Fig. 1 – Overview of the modeling analysis comparing different approaches to incorporating RPSFT in PSA. CE, cost-effectiveness; PSA, probabilistic sensitivity analysis; RPSFT, Rank Preserving Structural Failure Time.

ratio to receive either active therapy or standard care. Patients were entered into the analysis on the basis of sampling from a uniform random distribution over the course of 1 year. Underlying survival time for each patient was simulated from a Weibull distribution with parameters: shape 0.5, scale 1.33. Of the patients simulated, 30% were randomly selected to have "good" prognosis and had their survival inflated by a factor of 1.2. For patients in the standard care arm, treatment switching was applied to 10% and 25% for patients with "good" or "poor" prognosis, respectively. The time of switch was generated by sampling from a uniform distribution between study entry and death. An acceleration factor (ψ) of 1.23 (which corresponds to an HR of 0.9) was applied to patients in the active arm for the duration of their survival and to those in the standard care arm for the time from switching to the end of OS. A data cut-off time of 3 years was applied, with censoring implemented for any OS times extending beyond that period.

Assessment of Simulations

For each simulated data set, RPSFT was performed, a parametric model fit, and the associated covariance matrix was estimated by using the adjustment method and by bootstrapping. These estimated parameters and associated covariance matrices were then used as inputs into a simple economic model within R (R Core Team, 2014). This economic model was then used to evaluate overall (undiscounted) life expectancy from each of the 500 simulated data sets for both treatment arms. Statistical uncertainty around these estimates was estimated by performing PSA by sampling each of the derived variance/covariance matrices 5000 times.

Application of Test-Based Approach Using Published Summary Data

Individual patient data (IPD) is not commonly available to health service researchers, particularly for all relevant comparators, leaving those undertaking economic evaluations reliant upon published study summary results. A benefit of the test-based approach is the lack of reliance on acquiring the IPD when published unadjusted and RPSFT adjusted survival curves are available. An additional analysis was performed using the ITT and RPSFT-adjusted OS Kaplan-Meier (KM) estimates presented by Schlumberger et al. comparing lenvatinib to placebo for treatment of radioiodine-refractory thyroid cancer to demonstrate the approach [15].

The initial step was to recreate or simulate individual patient OS times from the KM data, using the algorithm by Guyot et al. [16]. Individual OS times were recreated from both the ITT or unadjusted and RPSFT-adjusted KM curves and numbers of patients at risk. To ensure the robustness of the recreated survival times, summary statistics from the publication and our analysis were compared (Supplementary Appendix Table 1).

Following the steps outlined in the earlier section, parametric Weibull functions were fit to the recreated survival times. The required adjustment factor, as outlined in equation 2, was then calculated before being applied to the covariance matrix (Supplementary Appendix Table 2). These estimated parameters and original and inflated covariance matrices were then used as inputs into a simple decision model within Microsoft Excel (Microsoft Corporation, Redmond, WA). The parameter estimates, along with the inflated covariance matrix were sampled 5000 times using PSA. The proportion of simulations where lenvatinib had a lower mean life expectancy compared with placebo was recorded.

Results

Mean estimates of OS and incremental life expectancy were similar when using all three methods across all 500 simulated data sets, as anticipated. However, the bootstrapped and test-based estimates consistently produced greater estimates of uncertainty compared with the conventional approach (Fig. 2). Using the conventional approach (not incorporating RPSFT model-attributed uncertainty into the variation around the parameter estimates), the width of the 95% CI for mean life years gained ranged from 0.20 to 0.29, with a mean of 0.24. When the bootstrap approach was used, this range increased to 0.20 to 0.36, with a mean of 0.26. The test-based adjustments to the variance/covariance matrix performed in a similar way to the bootstrap approach (Fig. 3). The similarity between the estimates from the bootstrapping approach and the proposed test-based approach indicate that the assumption made in the test-based approach—that the correlation between the parametric model parameters and treatment effect is not modified through the use of the RPSFT adjustment—is reasonable.

Similar results were observed using the test based approach using summary data; that is, failing to adjust the covariance matrix led to smaller CIs and suggested that the probability of placebo being determined to be more efficacious than lenvatinib was 0.74%, compared with 7.36% after the adjustment.

Discussion

In recent years, HTA decisions in oncology are increasingly being made on the basis of the results of RCTs that involve treatment switching. To support this process, decision analytic models for oncologic interventions are required to estimate life expectancy (and quality-adjusted life expectancy) by parametrically extrapolating survival times derived using the RPSFT method. HTA agencies commonly require that the results of PSA are presented alongside deterministic outcomes from cost-effectiveness models, to better understand the statistical uncertainty around
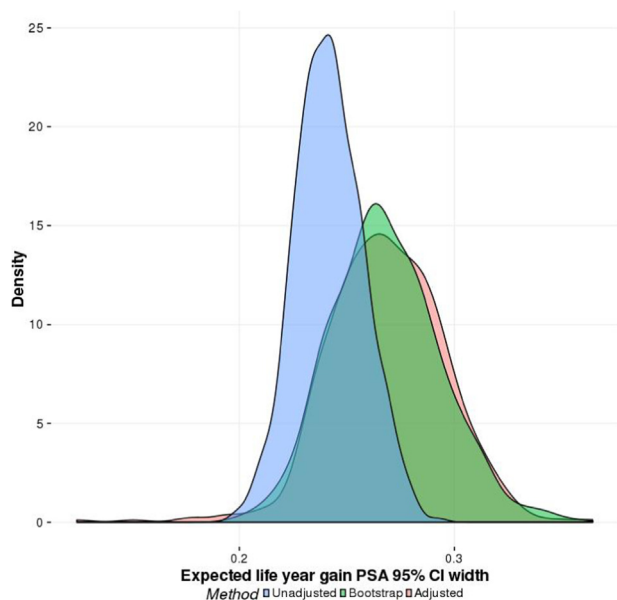

Fig. 2 – Illustration of results from the 500 simulations comparing different approaches to incorporating RPSFT model in PSA. PSA, probabilistic sensitivity analysis; RPSFT, Rank Preserving Structural Failure Time.
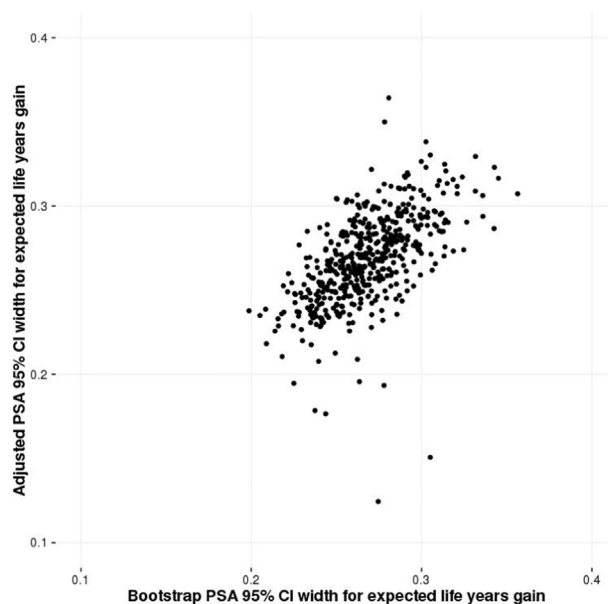

Fig. 3 – Illustration of results from the 500 simulations comparing different approaches to incorporating RPSFT model in PSA. PSA, probabilistic sensitivity analysis; RPSFT, Rank Preserving Structural Failure Time.

modeled outcomes. Typically, PSA is utilized to present the likelihood of an intervention being considered cost-effective by incorporating uncertainty around a range of model parameters. Conventionally, when parametric survival modeling is used, the variance/covariance matrix is typically sampled to propagate a distribution around estimates of life expectancy for both interventions. However, this approach does not account for the increased uncertainty in survival times attributed to the RPSFT method. The present study provides evidence that failure to accurately account for the increased uncertainty around life expectancy estimates associated with an RPSFT model can substantially influence the outcomes of PSA. Moreover, the current analysis only included estimates of life expectancy; had utility values and direct costs been included, the impact on the probability of a treatment being considered cost-effective may have been even greater.

An advantage of using an RPSFT model is that it creates a counterfactual data set (data that we would have observed had treatment switching not occurred) and has been shown to perform well across a range of simulation studies [14,17,18]. Standard parametric survival models used in decision models to estimate mean survival (and therefore life years gained, quality-adjusted life years gained, etc.) can be fitted to this counterfactual data set. The present study shows that without correctly adjusting for uncertainty around parameter estimates, there is a risk of reaching false conclusions (e.g., standard errors are lower and CIs are narrower, which may imply statistical significance that is not genuine and underestimating the costs of later lines of therapy). The mean treatment effect (HR, in the present example) should also be tested using the P value from the ITT analysis, rather than the treatment effect produced by Cox or parametric modeling using the counterfactual dataset.

Life expectancy in the decision analytic models were calculated using Weibull distribution for illustrative purposes, as this is one of the more common parametric models used in HTA submissions [19]. The principles described are equally applicable to other parametric functions, including mixture cure rate

models, which are increasingly utilized in the evaluation of immune-oncology treatment options [20].

Adjusting for treatment switching when results are confounded is likely to become an increasingly important aspect of HTA decision making in oncology in the years ahead because trials have reduced initial follow-up and use intermediate endpoints. The present analysis describes a novel approach to adequately adjust for uncertainty when using an RPSFT model, by simulating life expectancy in a hypothetical trial involving treatment switching. In addition, this analysis demonstrates the utility of the test-based approach in the case where IPD are not available preventing the use of bootstrapping. The evidence indicates that health researchers implementing treatment switching–adjusted survival data into decision models need to ensure uncertainty is handled appropriately to guard against spurious conclusions about the relative effectiveness (and therefore cost-effectiveness) of new interventions. Both the bootstrapping approach and the test-based approach provide a solution to appropriately incorporate uncertainty, with the benefit that the latter can implemented by researchers in the absence of IPD.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2017.07.008 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Motzer RJ, Escudier B, Oudard S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomized, placebo-controlled phase III trial. Lancet 2008;372:449–56.

[2] Latimer NR. Treatment switching in oncology trials and the acceptability of adjustment methods. Expert Rev Pharmacoecon Outcomes Res 2015;15:561–4.

[3] Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011;2:109–12.

[4] Latimer NR, Abrams KR, Lambert PC, et al. Adjusting survival time estimates to account for treatment switching in randomized controlled trials—an economic evaluation context: methods, limitations and recommendations. Med Decis Making 2014;34:387–402.

[5] Latimer N, Abrams K. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching, Report by the Decision Support Unit. July 2014. Available from: www.nicedsu.org.uk/TSD16_Treatment_Switching.pdf. [Accessed February 1, 2016].

[6] Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. Commun Stat Theory Methods 1991;20:2609–31.

[7] 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;56:779–88.

[8] Latimer N, Abrams K, Lambert PC, et al. Adjusting for treatment switching in randomized controlled trials—a simulation study and a simplified two-stage method. Stat Methods Med Res 2017;26:724–51.

[9] NICE. Final appraisal determination: vemurafenib for treating locally advanced or metastatic BRAF V600 mutation-positive malignant melanoma. 2012. Available from: www.nice.org.uk/guidance/ta269/documents/melanoma-braf-v600-mutation-positive-unresectable-metastatic-vemurafenib-final-appraisal-determination-document2. [Accessed January 26, 2016].

[10] Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14:339–47.

[11] Baio G, Dawd AP. Probabilistic Sensitivity Analysis in Health Economics. Research Report No. 292, Department of Statistical Science, University College London. March, 2008. Available from: https://www.ucl.ac.uk/statistics/research/pdfs/rr292.pdf. [Accessed January 27, 2016].

[12] White IR, Babiker AG, Walker S, Darbyshire JH. Randomization-based methods for correcting for treatment changes: examples from The Concorde Trial. Stat Med 1999;18:2617–34.

[13] Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford, UK: Oxford University Press, 2006.

[14] Morden JP, Lambert PC, Latimer N, et al. Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. BMC Med Res Methodol 2011;11:1–20.

[15] Schlumberger M, Tahara M, Wirth LJ, et al. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. N Engl J Med 2015;372:621–30.

[16] Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012;12:9.

[17] Latimer NR, 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;26:724–51.

[18] Latimer NR, Abrams KR, Lambert PC, et al. Assessing methods for dealing with treatment switching in clinical trials: a follow-up simulation study. Stat Methods Med Res 2016. http://dx.doi.org/10.1177/0962280216642264.

[19] Latimer N. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data. Inconsistencies, limitations and a practical guide. Med Decis Making 2013;33:743–54.

[20] Hellman MD, Ma J, Garon EB, et al. Estimating long-term survival of PD-L1-expressing, previously treated, non-small cell lung cancer patients who received pembrolizumab in KEYNOTE-001 and -010. J Clin Oncol 2017:35: Abstract 77.