Adjusting overall survival for treatment switch

Recommendations of a cross-institutional statistical working group

Claire Watkins

BBS/EFSPI European Scientific Meeting
Application of Methods for Health Technology Assessment
23rd June 2015

Disclosure statement: Claire Watkins is an employee of AstraZeneca UK Ltd. The views and opinions expressed herein are my own and cannot and should not necessarily be construed to represent those of AstraZeneca or its affiliates.
Outline

The working group

Background

Methods

Assumptions and limitations

Example

Best practice – design and analysis
Treatment switch subteam of the PSI HTA SIG

Remit and membership

Amongst pharmaceutical industry statisticians working with trials including treatment switches:

- Raise awareness
  - Current statistical methods
  - Potential applications
  - Strengths and limitations
- Promote and share best practice, in both analysis and design
- Encourage research into new and improved methods

[Chair] Claire Watkins (AstraZeneca)
Pierre Ducournau (Roche)
Rachel Hodge (GSK)
Xin Huang (Pfizer)
Cedric Revil (Roche)

Alexandros Sagkriotis (Novartis)
Yiyun Tang (Pfizer)
Elaine Wright (Roche)
[Non-member advisory role]
Nicholas Latimer (University of Sheffield)
Adjusting overall survival for treatment switches: commonly used methods and practical application

Claire Watkins, Xin Huang, Nicholas Latimer, Yiyun Tang, and Elaine J. Wright

In parallel group trials, long-term efficacy endpoints may be affected if some patients switch or cross over to the alternative treatment arm prior to the event. In oncology trials, switch to the experimental treatment can occur in the control arm following disease progression and potentially impact overall survival. It may be a clinically relevant question to estimate the efficacy that would have been observed if no patients had switched, for example, to estimate ‘real-life’ clinical effectiveness for a health technology assessment. Several commonly used statistical methods are available that try to adjust time-to-event data to account for treatment switching, ranging from naive exclusion and censoring approaches to more complex inverse probability of censoring weighting and rank-preserving structural failure time models. These are described, along with their key assumptions, strengths, and limitations. Best practice guidance is provided for both trial design and analysis when switching is anticipated. Available statistical software is summarized, and examples are provided of the application of these methods in health technology assessments of oncology trials. Key considerations include having a clearly articulated rationale and research question and a well-designed trial with sufficient good quality data collection to enable robust statistical analysis. No analysis method is universally suitable in all situations, and each makes strong untestable assumptions. There is a need for further research into new or improved techniques. This information should aid statisticians and their colleagues to improve the design and analysis of clinical trials where treatment switch is anticipated. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: treatment switching; crossover; inverse probability of censoring weighting; rank-preserving structural failure time; health technology assessment
Background

What do we mean by treatment switch/crossover?

Patients in a parallel group RCT may switch or “crossover” to the alternative treatment at some point before an endpoint of interest occurs.

- We might build this into the trial protocol
  - e.g. Sunitinib GIST trial (Demetri, 2012)

- Or it might happen spontaneously due to clinical practice in the region, if the treatment is already on the market
  - e.g. Gefitinib IPASS trial (Fukuoka, 2011)
Potential impact of treatment switch
Correlation between PFS and OS in NSCLC


Switch prohibited (n=20)
R²=0.5341

Switch allowed (n=15)
R²=0.0027

Interaction p=0.019
Why does this matter for HTA?
It all depends on the decision problem

In HTA, the decision problem is often to compare:

Current clinical practice **without** new therapy  vs  Potential future clinical practice **including** new therapy

If there is switching and the new therapy is effective, ITT underestimates this difference

→ *How to estimate long term efficacy without switch?*
Commonly used methods to estimate control arm survival in absence of switch

“Naive” methods
• Exclude switchers
• Censor at switch
• Time varying covariate

Simple to apply
High levels of bias

“Complex” methods
• Inverse Probability of Censoring Weighting (IPCW; observational)
• Rank Preserving Structural Failure Time (RPSFT; randomisation based)
• Two-stage Accelerated Failure Time
• Others

Harder to apply
Try to reduce bias

External data
Naive (1): Exclude switchers

Control arm survival
Compare to observed experimental arm survival

Key
- Death time
- Censor time
- Switch time

Observed (ITT) control arm

Exclude switchers

Assumption: Switchers and non-switchers have the same prognosis
Naive (2): Censor switchers
Control arm survival
Compare to observed experimental arm survival

Key
- Death time
> Censor time
S Switch time

Assumption: Switchers and non-switchers have the same prognosis
Naive (3): Time varying covariate

Assumption: Switchers and non-switchers (with the same covariates) have the same prognosis
(i.e. no confounders)
**Complex (1): IPCW (weight non-switched times)**

Control arm survival

Compare to observed experimental arm survival

---

**Observed (ITT) control arm**

- Non switchers
- Switchers

**IPC weighted**

- Non switchers
- Switchers

Key

- Death time
- Censor time
- Switch time

**Assumption:** The variables in the weight calculation fully capture all reasons for switching that are also linked to survival (i.e. no unmeasured confounders)

Weights represent how “switch-like” a patient is that has not yet switched

Calculated using observational propensity score methods, vary by time and patient
**Statistical detail: IPCW**
An observational data-based two stage modelling approach

**Modelling stage 1:**

<table>
<thead>
<tr>
<th>Population</th>
<th>Control arm patients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Probability of switch over time conditional on baseline and time varying factors (propensity score)</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>Time varying subject specific IPC weights</td>
</tr>
</tbody>
</table>

**Modelling stage 2:**

<table>
<thead>
<tr>
<th>Population</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Survival analysis with IPC weights applied to control arm patients (censor at switch)</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td>IPC weighted Hazard Ratio and KM</td>
</tr>
</tbody>
</table>
Statistical detail: IPCW
Modelling stage 1: Obtaining IPC weights

Time dependent Cox model for switch (Robins 2000)

$$\lambda_C(t|\bar{V}(t), T>t) = \lambda_0(t) \exp{\alpha'V(t)}$$

Core assumption: $V$ includes all covariates that influence both the decision to switch and survival
If not: Estimate will be biased
Implication: Additional data collection
Statistical detail: IPCW
Modelling stage 1: Obtaining IPC weights

Usually fit using pooled logistic regression (Hernan 2000, Fewell 2004)
- Analogous to time dependent Cox (D’Agostino, 1990)
- Split data into patient cycles and use as observations
- Time varying covariates set as constant within a patient-cycle
- Delete any cycles after switch
- Intercepts via a spline function
- Use patient clustering to ensure robust SE incorporating within patient correlation (e.g. SAS REPEATED statement)

\[
\text{Logit } P(\text{switch in cycle } k) = \beta' V(k)
\]

IPC weight for cycle k

\[
\prod_{j=1..k} P(\text{do not switch in cycle } j| V_{\text{base}}(j)) \cdot P(\text{do not switch in cycle } j| V(j))
\]
Statistical detail: IPCW

Modelling stage 2: IPC weighted survival analysis

Experimental arm: Observed data (weight 1)
Control arm prior to switch: IPC weights
Control arm after switch: Censor (zero weight)

Often fitted using pooled logistic regression as per stage 1
(constant weight within a patient cycle)

IPC weighted control arm Kaplan Meier curve can be generated with some effort (Robins 2000)
Complex (2): RPSFT (adjust post switch times)
Control arm survival
Compare to observed experimental arm survival

Assumptions: Each cycle of treatment extends survival by a constant amount. Balanced arms due to randomisation.
Statistical detail: RPSFTM (Robins 1991)  
Accelerated Failure Time model structure

Works by estimating the treatment effect that balances “counterfactual” survival time in the absence of treatment \([T(0)]\) between randomised arms.

\[ T(0) = T_{\text{off}} + e^\psi T_{\text{on}} \]

Survival time without experimental  
Time off experimental  
Treatment effect  
Time on experimental

Treatment effect <1 is beneficial (time to death is slowed down on trt)  
Treatment effect >1 is detrimental (time to death is speeded up on trt)

Core assumption: Trt effect is the same regardless of when treated

NOTE: RPSFTM does not increase power -> same p-value as ITT
Statistical detail: RPSFTM

Estimating $\psi$ : The G-estimation procedure

Due to randomisation, assume $T(0)$ distribution is the same for both arms

For a range of $\psi$ values, calculate $T(0)$ and test if equal (using a standard survival analysis test statistic $Z$)

Select $\psi$ that best satisfies test statistic $Z(\psi) = 0$

Core assumption: randomised arms are balanced
Statistical detail: RPSFTM

RPSFT adjusted survival analysis

Use observed survival time and event flag for experimental and control arm non-switch patients

Use $T(0) = \text{estimated survival time in absence of exposure to experimental for control arm switch pts}$
  - Re-censoring should be applied to event flag

Perform standard survival analyses to get the HR & KM

SE from standard analysis too small – does not incorporate uncertainty in $\psi$
  - Bootstrap CI, or set p-value=ITT and derive CI limits accordingly
RPSFTM (not) in SAS

Vote in the SASware ballot for this to be added as a standard analysis option:

https://communities.sas.com/ideas/1325
Complex (3): 2-Stage AFT (observational study)

Control arm survival
Compare to observed experimental arm survival

Observed (ITT) control arm

Non switchers

Switchers

2-stage AFT adjusted

- Treat control arm as observational study post progression
- Re-baseline at progression
- Collect covariate data at progression
- Calculate effect of switch treatment adjusting for covariates
- Adjust switcher data and compare randomised arms

Assumptions: The covariates fully capture all reasons for switching that are also linked to survival
No time-dependent confounding between P and S
(i.e. no unmeasured confounders)

Not valid if switch can occur before progression
Does not require covariate data collection after progression

Key
- Death time
- Censor time
- Switch time
- Progression time
Use of external data

Control arm without exposure to experimental

Use as external validation of analyses, or quasi-control

Comparability to pivotal RCT control arm is key - Patient characteristics, eligibility criteria, treatment, time, location, clinical practice, outcome
### Key assumptions, strengths and limitations

**“Naive” methods**

<table>
<thead>
<tr>
<th></th>
<th>Key Assumptions</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT</td>
<td>Switch does not affect survival</td>
<td>As per RCT design</td>
<td>Assumption unlikely to hold, leading to <strong>bias</strong></td>
</tr>
</tbody>
</table>
Key assumptions, strengths and limitations
“Complex” methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Key Assumptions</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPSFTM (Rank Preserving</td>
<td>Constant treatment effect. Counterfactual survival time is balanced between</td>
<td>Reduced selection bias. Performs well if constant treatment effect.</td>
<td>Bias if treatment effect not constant (disease specific). Poor performance if</td>
</tr>
<tr>
<td>Structural Failure Time</td>
<td>treatment groups due to randomization.</td>
<td>preserves randomisation.</td>
<td>most patients switch or near perfect switch predictor. Can recover lost</td>
</tr>
<tr>
<td>Model)</td>
<td></td>
<td></td>
<td>power due to switch (anti-conservative) or lose power due to loss of events.</td>
</tr>
<tr>
<td>IPCW (Inverse Probability of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Censoring Weighting)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Overall Survival Probability (%)

Sunitinib (N=243)
Median 72.7 weeks
95% CI (61.3, 83.0)

Placebo (N=118)
Median 64.9 weeks
95% CI (45.7, 96.0)

Hazard Ratio=0.876
95% CI (0.679, 1.129)
p=0.306

103 (87.3%) patients crossed over from placebo to sunitinib treatment
### Conventional Analyses

<table>
<thead>
<tr>
<th>Cox Model</th>
<th>Hazard Ratio (SU/PB), 95% CI and p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITT (naïve)</td>
<td>0.876 (0.679, 1.129), p=0.306</td>
</tr>
<tr>
<td>Dropping crossover</td>
<td>0.315 (0.178, 0.555), p&lt;0.0001</td>
</tr>
<tr>
<td>Censoring at crossover</td>
<td>0.825 (0.454, 1.499), p=0.527</td>
</tr>
<tr>
<td>Time-dependent treatment</td>
<td>0.934 (0.520, 1.679), p=0.820</td>
</tr>
</tbody>
</table>
Overall Survival (Final, 2008)
Crossover Adjusted by RPSFT

Overall Survival Probability (%)
Time (Week)

Sunitinib (N=243)
Median 72.7 weeks
95% CI (61.3, 83.0)

Placebo (N=118)
Median* 39.0 weeks
95% CI (28.0, 54.1)

Hazard Ratio=0.505
95% CI** (0.262, 1.134)
p=0.306

Sunitinib (N=207)
Placebo (N=105)

*Estimated by RPSFT model
**Empirical 95% CI obtained using bootstrap samples.
Best practice recommendations - Trial design

General

- Up front planning at design stage
- Be clear about the question, rationale, customer
- Describe intent to do analyses in protocol
- Define switch treatment (drug(s), dose(s) etc)
- Consider external data sources
- Collect the data to enable robust analysis
When (not) to build in switch

• Generally preferable not to

• Situations where rationale more compelling:
  
  Strong efficacy signal from existing data
  
  Spontaneous switch likely, wish to manage
  
  Recommended by external bodies
  
  Trial becomes infeasible without
  
  Primary question is sequencing
Best practice recommendations – Trial design
Additional requirements with Built in switch

- Clear and unambiguous criteria for switch
- Blinded evaluation of criteria
- Identical assessments for each arm
Best practice recommendations - Analysis

- Pre-specify preferred adjustment method in protocol/analysis plan with justification
- Assess plausibility of assumptions
- Don’t use naive as sole or primary method
- Don’t use IPCW/2-stage AFT if most patients switch or near perfect predictor of switch
- Don’t use RPSFTM if near equal exposure or survival
- Always present ITT
- Provide results from a range of sensitivity analyses
- External data – assess and take steps to maximise comparability

Take care when extrapolating switch-adjusted data!
Summary and looking to the future

- No method is universally “best”
  - Situation specific assessment required

- Best practice recommendations have been outlined
  - Be clear on the decision problem
  - Collect the right data
  - Up front planning

- Further work and research is needed:
  - Multiple switches
  - Multiple switch treatments
  - Different or less strong assumptions
  - Binary or continuous data
  - Software
References (1)


Hernan MA et al. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000; 11(5):561-570


Huang X and Xu Q. Adjusting the Crossover Effect in Survival Analysis Using a Rank Preserving Structural Failure Time Model: The Case of Sunitinib GIST Trial. MRC HTMR workshop, Feb 2012, available here
Backup slides
Statistical software

- Naive methods – standard survival analysis software

- IPCW (a simplified Marginal Structural Model)
  - SAS code for MSMs available in Hernan 2000 and at http://www.hsph.harvard.edu/causal/software/ (second example)
  - Stata code for MSMs available in Fewell 2004
  - No published code available for IPC weighted KMs

- RPSFTM
  - Stata strbee package available (White 2002)
Trial design features

Data collection (may require new CRFs)

For general understanding (optional):
• Built in: If and when switch criteria are met, reasons for not following criteria
• Spontaneous: Reason for switch
• Outcome of switch therapy

For RPSFTM and IPCW (required):
• Drug, dosing, etc (to identify switch therapy)
• Start and stop dates of switch therapy

For IPCW (required):
• Baseline characteristics that may influence decision to switch
• Time varying factors that may influence decision to switch
During ongoing study conduct

- Monitor switch therapy
- Built in switch: monitor visit dates vs schedule
- Take action if needed
  - E.g. amend protocol or analysis plan if original assumptions wrong
Take care with extrapolation of adjusted data
Latimer and Abrams 2014

<table>
<thead>
<tr>
<th>Extrapolation method</th>
<th>Switch adjustment method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RPSFTM 2-stage AFT</td>
</tr>
<tr>
<td>Fit separate parametric models to each treatment group</td>
<td>Fit to adjusted patient data. Re-censoring required</td>
</tr>
<tr>
<td>Fit one parametric model to all data and apply proportional treatment effect</td>
<td>Fit to adjusted patient data. Re-censoring required</td>
</tr>
</tbody>
</table>
Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. AstraZeneca PLC, 2 Kingdom Street, London, W2 6BD, UK, T: +44(0)20 7604 8000, F: +44 (0)20 7604 8151, www.astrazeneca.com