Methods to account for treatment switching in randomised controlled trials

Leeds, 14th October 2015

Ian White
MRC Biostatistics Unit, Cambridge, UK
Plan

1. Motivation: some recent trials with a big treatment switching problem
2. What’s the question?
   - and does intention-to-treat answer it?
   - will mainly talk about clinical effectiveness, but also consider implications for cost-effectiveness
3. What methods are available?
   - intention-to-treat
   - per-protocol
   - inverse-probability-of-censoring weighting (IPCW)
   - instrumental variable methods; rank-preserving structural nested failure time models (RPSFTMs)
4. Improving the RPSFTM
5. Implications for practice
Plan

1. Motivation: some recent trials with a big treatment switching problem
2. What’s the question?
3. What methods are available?
4. Improving the RPSFTM
5. Implications for practice
Motivation 1: Sunitinib trial

- RCT evaluating sunitinib for patients with advanced gastrointestinal stromal tumour after failure of imatinib

- Interim analysis found big treatment effect on progression-free survival

- All patients were then allowed to switch to open-label sunitinib

- Next slides are from Xin Huang (Pfizer)
Time to Tumor Progression
(Interim Analysis Based on IRC, 2005)

- Sunitinib (n=178)
  - Median: 6.3, 95% CI (3.7, 7.6)
- Placebo (n=93)
  - Median: 1.5, 95% CI (1.0, 2.3)

Hazard Ratio = 0.335
p < 0.00001

with thanks to Xin Huang (Pfizer)
Overall Survival (NDA, 2005)

Overall Survival Probability (%)

- **Sunitinib (N=207)**
  - Hazard Ratio = 0.49
  - 95% CI (0.29, 0.83)
  - p = 0.007

- **Placebo (N=105)**

Total deaths:
- 29 (Sunitinib)
- 27 (Placebo)

with thanks to Xin Huang (Pfizer)
Overall Survival (ASCO, 2006)

- **Sunitinib (N=243)**
  - Overall Survival Probability (%)
  - Hazard Ratio=0.76
  - 95% CI (0.54, 1.06)
  - p=0.107

- **Placebo (N=118)**
  - Overall Survival Probability (%)
  - Total deaths: 89 & 53

*with thanks to Xin Huang (Pfizer)*
Overall Survival (Final, 2008)

- **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

**Total deaths**
- Sunitinib: 176
- Placebo: 90

with thanks to Xin Huang (Pfizer)
Sunintinib: explanation?

- The decay of the treatment effect is probably due to treatment switching
- Of 118 patients randomized to placebo arm:
  - 103 patients switched to sunitinib treatment
    - 83 switched within 3 months
    - 19 switched before disease progression
    - 4 never treated with placebo
  - 15 patients did not switch
Motivation 2: Everolimus trial

- “Treatment in both groups was continued until disease progression, unacceptable toxicity, death, or discontinuation for any other reason. Investigators were unaware of the study group assignments, but disclosure was permitted after documented progression on the basis of investigator assessment. Patients who were initially randomised to placebo were then able to crossover to receive open-label everolimus. This element of the study design was incorporated to address both ethical and recruitment considerations.”
Everolimus: progression-free survival

Hazard ratio 0.30, 95% CI 0.22-0.40; p<0.0001

Everolimus: overall survival

Hazard ratio 0.83, 95% CI 0.50-1.37; p=0.23
Everolimus: overall survival

• “There was no significant difference between groups in terms of overall survival (hazard ratio 0.83, 95% CI 0.50-1.37; p=0.23), probably due to confounding by crossover:
  – of the 98 patients in the placebo group who progressed as per investigator assessment, 79 crossed over to open-label everolimus after disease progression.
  – 60 of these 79 patients had progressed within 8 weeks of enrolment.”
Why does it matter?

**Drug licensing**
- Clear benefit on progression-free survival could be enough for drug approval by regulators
- May be OK to ignore overall survival hence sidestep the switching issue

**Funding decisions / cost-effectiveness analysis / NICE**
- Must take account of overall survival
NICE guidance

  - manufacturer used RPSFTM analysis for overall survival to correct for switches
  - approved
  - manufacturer used IPCW analysis for overall survival to correct for switches
  - NICE asked for RPSFTM to be done too: slightly less favourable results
  - not approved
Plan

1. Motivation: some recent trials with a big treatment switching problem
2. What’s the question?
3. What methods are available?
4. Improving the RPSFTM
5. Implications for practice

Statistician: "What's the estimand?"
Health economist: "What's the decision problem?"
Defining the question

- For sunitinib, the main question was
  - “drug now”: treatment as actually given in the sunitinib arm (given until clinical decision to stop, usually due to adverse event / progression)
  vs.
  - “no drug”: no drug at all, even after progression, because it hasn’t been approved
- Instead the trial answered
  - “drug now”: as actually given in sunitinib arm
  - vs. “deferred drug”: as actually given in placebo arm
- That is, the RCT didn't address the main question
- This is a common, but not universal, setting
Part of a wider problem

- Note on terminology: people often talk about “treatment cross-overs”
  - to avoid confusion with cross-over trials, I use “treatment switches”
- Many trials have not just treatment switching (i.e. to the treatment allocated to the other trial arm), but also more general “non-compliance” (term from the statistical literature):
  - other changes of prescribed treatment
    » changes to non-trial treatments
    » changes to no treatment
    » multiple treatments
    » dose adjustment
  - non-compliance with prescribed treatment
Three common questions

• What is the effect of assignment to treatment A *in the circumstances of the trial?* (effectiveness; de facto)
  – could be: A immediately vs. A on progression
  – intention-to-treat analysis answers this question

• What will be the effect of assignment to treatment A *in other circumstances?* (alternative effectiveness)
  – could be: A immediately (for as long as tolerated) vs. no A
  – sunitinib example: NICE’s question was sunitinib immediately (with discontinuations as in clinical practice) vs. no sunitinib

• What is the effect of treatment A *per se* (efficacy; de jure)?
  – i.e. while actually given
Intention-To-Treat (ITT) Analysis

- Compares outcomes for participants as randomised
  - ignores treatment actually received
- Evaluates the effect of assignment to treatment rather than the effect of treatment receipt
  - needs fewer assumptions than other methods
- Essential part of analysis
  - an unbiased answer, but only to the "effectiveness" question
Is ITT analysis adequate?

- If our question of interest is “What is the effect of assignment to treatment A in the circumstances of the trial?” (effectiveness) then ITT analysis is adequate
- But still need to report switches so that readers know what “the circumstances of the trial” are
- A review of 100 recent trials (Dodd et al) found that “Nonadherence to treatment protocol was reported in 98 of the 100 trials, but ...
  - “42 publications did not state how many participants started their randomised treatment.
  - “Reporting of treatment initiation and completeness was judged to be inadequate in 64% of trials with short-term interventions and 89% of trials with long-term interventions.”

Defining the question: counterfactuals often help

<table>
<thead>
<tr>
<th>Counterfactual</th>
<th>Setting</th>
<th>Possible estimand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (what patient (i) would have received if they had been randomised to treatment A)</td>
<td>Binary treatment</td>
<td>difference between arms in the subgroup who would take treatment if randomised to it (\rightarrow) complier average causal effect, CACE</td>
</tr>
<tr>
<td>Outcome (outcome for patient (i) if they had received treatment A; or if they had been randomised to treatment A)</td>
<td>Any</td>
<td>difference between arms if there had been no departures [of a particular type] from randomised treatment (\rightarrow) main focus here</td>
</tr>
</tbody>
</table>
Plan

1. Motivation: some recent trials with a big treatment switching problem
2. What’s the question?
3. What methods are available?
4. Improving the RPSFTM
5. Implications for practice
Statistical methods to handle switching

- Intention-to-treat analysis
- Per-protocol analysis
- Inverse-probability-of-censoring weighting (IPCW)
- Instrumental variables (IVs)
- Rank-preserving structural nested failure time models (RPSFTMs)

I will assume we are analysing
- individual-level data (but one can make some progress using just trial reports)
- from a randomised trial (but observational data can be used to mimic a RCT – still need to allow for switching)
Hypothetical trial data (observed counts)

Randomise: drug A or placebo

1\textsuperscript{st} follow-up: detect disease progression; no deaths at this stage

2\textsuperscript{nd} follow-up: look at mortality

<table>
<thead>
<tr>
<th>Arm</th>
<th>Time 1 Progression</th>
<th>Switch Progression</th>
<th>Time 2 status Dead/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>No (800)</td>
<td>No (800)</td>
<td>10 / 800</td>
</tr>
<tr>
<td></td>
<td>Yes (200)</td>
<td>Yes (200)</td>
<td>90 / 200</td>
</tr>
<tr>
<td>Placebo</td>
<td>No (600)</td>
<td>No (600)</td>
<td>10 / 600</td>
</tr>
<tr>
<td></td>
<td>Yes (400)</td>
<td>Yes (200)</td>
<td>60 / 200</td>
</tr>
</tbody>
</table>

Placebo arm progressors may switch to Drug A

Question: what would the difference between the two arms be if no switching occurred in the placebo arm?
Hypothetical trial: ITT analysis

<table>
<thead>
<tr>
<th>Arm</th>
<th>Time 1 Progression</th>
<th>Switch</th>
<th>Time 2 status Dead/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>No (800)</td>
<td></td>
<td>10 / 800</td>
</tr>
<tr>
<td></td>
<td>Yes (200)</td>
<td></td>
<td>90 / 200</td>
</tr>
<tr>
<td>Placebo</td>
<td>No (600)</td>
<td>No (600)</td>
<td>10 / 600</td>
</tr>
<tr>
<td></td>
<td>Yes (400)</td>
<td>No (200)</td>
<td>90 / 200</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes (200)</td>
<td>60 / 200</td>
</tr>
</tbody>
</table>

Result: **100/1000 vs. 160/1000**
Per-protocol (PP) Analysis

- Censors participants who switch from their randomly-allocated treatments (at the time of switch)
- Hence not based on everyone as randomised
- Subject to possible selection biases (confounding)
  - prognosis likely to be different in those who switch treatments (e.g. they may be sicker)
  - selection bias can be reduced by using IPCW (next)

- Despite its potential disadvantages, per-protocol analysis is often advocated alongside ITT in the analysis of non-inferiority trials
Hypothetical trial: per-protocol analysis

<table>
<thead>
<tr>
<th>Arm</th>
<th>Progression</th>
<th>Switch</th>
<th>Time 2 status Dead/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>No (800)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10 / 800</td>
</tr>
<tr>
<td></td>
<td>Yes (200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>90 / 200</td>
</tr>
<tr>
<td>Placebo</td>
<td>No (600)</td>
<td>No (600)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10 / 600</td>
</tr>
<tr>
<td></td>
<td>Yes (400)</td>
<td>No (200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>90 / 200</td>
</tr>
<tr>
<td></td>
<td>Yes (200)</td>
<td>Yes (200)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60 / 200</td>
</tr>
</tbody>
</table>

Result: **100/1000** vs. **100/800**
Inverse-probability-of-censoring weighting (IPCW) methods

- Like per-protocol analysis, IPCW artificially censors a patient’s follow-up when they switch treatment.
- To tackle selection bias, censored (switched) patients are taken into account by increasing the weight of similar uncensored (unswitched) patients when fitting the “outcome model” (e.g. Cox model for death on randomised group).
  - weight = 1 / P(unswitched | covariates)
  - requires a “switching model” to predict censoring (switching)
  - to avoid bias, switching model must include all baseline and time-dependent variables that predict both treatment switching and outcome: “no unmeasured confounders”

# IPCW analysis

<table>
<thead>
<tr>
<th>Arm</th>
<th>Time 1 Progression</th>
<th>Switch</th>
<th>Time 2 status Dead/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>No (800)</td>
<td></td>
<td>10 / 800 1</td>
</tr>
<tr>
<td></td>
<td>Yes (200)</td>
<td></td>
<td>90 / 200 1</td>
</tr>
<tr>
<td>Placebo</td>
<td>No (600)</td>
<td>No (600)</td>
<td>10 / 600 1</td>
</tr>
<tr>
<td></td>
<td>Yes (400)</td>
<td>No (200)</td>
<td>90 / 200 2</td>
</tr>
<tr>
<td></td>
<td>Yes (200)</td>
<td></td>
<td>60 / 200 0</td>
</tr>
</tbody>
</table>

\[ P(\text{switch} \mid \text{placebo, progressed}) = 0.5 \rightarrow \text{non-switchers get weight 2} \]

Assumes switchers & non-switchers have comparable counterfactual switch-free outcomes

Result: 100/1000 vs. 190/1000
Constructing IPCW weights in general

- Need a model for switching given baseline and time-dependent covariates $X_{it}$
- Switching models:
  - discrete time: logit $P(i$ switches at $t) = \alpha_t + \beta'X_{it}$
  - continuous time – Cox: $h(t) = h_0(t) \exp(\beta'X_{it})$
- Fit switching model & hence estimate $p_{it} = P($individual i hasn't switched by time t$)$ for all outcome-event times $t$
- Weight the analysis of the outcome model by $w_{it} = 1/p_{it}$
  - need robust (sandwich) standard errors, or bootstrapping, to allow for the weighting
  - time-dependent weights are a problem in some software: Fewell et al (2004) recommend using pooled logistic regression for discrete times in Stata

Choice of covariates for IPCW

- Recall: anything that predicts both switching and outcome
- Baseline covariates: the usual stuff
- Time-dependent covariates (very important):
  - disease progression (as in Sunitinib and Everolimus examples)
  - disease severity (performance status etc.)
  - any other prognostic variable that clinicians would use to decide whether to switch
The problem of unstable weights

- Sometimes we get very large weights in IPCW
  - e.g. if 99% of patients who progressed then switched, the 1% who didn’t switch get a weight of 100 to “represent” those who did switch
- Leads to large standard errors
- “Capping” weights avoids large standard errors
  - but re-introduces bias
- “Stabilised” weights help if baseline covariates are strong predictors of switch (Robins et al, 2000)
- This is an inherent limitation of the method
  - e.g. IPCW must fail if 100% of patients who progressed then switched

Final note on IPCW

• IPCW can handle more than just switching:
  – any other sort of treatment changes
  – other “protocol violations” such as loss to follow up
• The “no unmeasured confounders” assumption must be assessed using knowledge about the trial and the data
Instrumental variables (IV)

- Standard IV methods are useful for a quantitative outcome (e.g. costs or QoL)
- Basic idea: randomisation is the IV
- Key assumption: randomisation only affects outcome via treatment ("exclusion restriction")
- In the very simplest case (no covariates, all-or-nothing treatment) we can say
  - ITT difference = effect of treatment * \{p(\text{treat} \mid \text{rand=treat}) - p(\text{treat} \mid \text{rand=control})\}
  - hence effect of treatment = ITT difference / \{p(\text{treat} \mid \text{rand=treat}) - p(\text{treat} \mid \text{rand=control})\}
  - Stata `ivregress 2sls effect (treat = rand)`

```bash
Stata ivregress 2sls effect (treat = rand)
```
Instrumental variables (IV) ctd.

- In the case of treatment effect heterogeneity, the quantity estimated by IV turns out to be the average treatment effect among the “compliers” – those who would be treated if and only if randomised to treatment
  - Complier Average Causal Effect (Angrist et al 1996)
- For a survival outcome, things get much more complicated
  - unfortunately that’s what we need to tackle
- I’ll describe the method which extends IV ideas for survival outcomes ...

Rank-preserving structural failure time model (1)

• Observed data for individual $i$:
  - $Z_i$ = randomised group
  - $D_i(t)$ = whether on treatment at time $t$
  - $T_i$ = observed outcome (time to event)

• Ignore censoring for now

• The RPSFTM relates $T_i$ to a potential outcome $T_i(0)$ that would have been observed without treatment through a treatment effect $\psi$ (Robins JM, Tsiatis AA. Comm Stats Theory Meth 1991; 20(8): 2609–2631)

• Case 1: all-or-nothing treatment (e.g. surgery)
  - treatment multiplies lifetime by a ratio $\exp(-\psi)$
  - $\psi < 0$ means treatment is good
  - untreated individuals: $T_i = T_i(0)$
  - treated individuals: $T_i = \exp(-\psi) \times T_i(0)$
Rank-preserving structural failure time model (2)

- Case 2: time-dependent 0/1 treatment (e.g. drug prescription, ignoring actual adherence)
  - define $T_i^{\text{off}}$, $T_i^{\text{on}}$ as follow-up times off and on treatment
    - so $T_i^{\text{off}} + T_i^{\text{on}} = T_i$
  - treatment multiplies just the $T_i^{\text{on}}$ part of the lifetime
  - model: $T_i(0) = T_i^{\text{off}} + \exp(\psi) \times T_i^{\text{on}}$

- General model handles time-dependent quantitative treatment (e.g. drug adherence):
  \[ T_i(0) = \int_0^{T_i} \exp\{\psi D_i(t)\} \, dt \]

- Interpretation: your assigned lifetime $T_i(0)$ is used up $\exp(\psi)$ times faster when you are on treatment
  - $\exp(\psi)$ is the acceleration factor
RPSFTM: identifying assumptions

Model: $T_i(0) = T_i^{off} + \exp(\psi) \times T_i^{on}$

- Common treatment effect
  - treatment effect, expressed as $\psi$, is the same for both arms
  - strong assumption if the control arm is (mostly) treated from progression while the experimental arm is treated from randomisation
  - want to do sensitivity analyses → Improvement 1

- Exclusion restriction
  - untreated outcome $T(0)$ is independent of randomised group $Z$
  - usually very plausible in a double-blind trial

- Comparability of switchers & non-switchers is NOT assumed
G-estimation: an unusual estimation procedure

Model: \( T_i(0) = T_i^{off} + \exp(\psi) \times T_i^{on} \)

- Take a range of possible values of \( \psi \)
- For each value of \( \psi \), work out \( T(0) \) and test whether it is balanced across randomised groups
- Graph test statistic against \( \psi \)
- Best estimate of \( \psi \) is where you get best balance (smallest test statistic)
- 95% CI is values of \( \psi \) where test doesn’t reject
- User has free choice of test
- Conventionally the same test as in the ITT analysis
  - typically log rank test \( \rightarrow \) Improvement 2
RPSFTM: P-value

Model: \( T_i(0) = T_i^{off} + \exp(\psi) \times T_i^{on} \)

- When \( \psi = 0 \) we have \( T_i(0) = T_i \)
- So the test statistic is the same as for the observed data
- Thus the P-value for the RPSFTM is the same as for the ITT analysis (provided the same test is used for both)
  - logic: null hypotheses are the same
  - under the RPSFTM, \( T \perp Z \) if and only if \( \psi = 0 \)
- The estimation procedure is “randomisation-respecting”
  - it is based only on the comparison of groups as randomised
RPSFTM: Censoring

- Censoring introduces complications in RPSFTM estimation
  - censoring on the $T(0)$ scale is informative
  - requires re-censoring which can lead to strange results

Estimating a causal hazard ratio

- Often hard to interpret $\psi$
- Use the RPSFTM again to estimate the untreated event times $T_i(0)$ in the placebo arm
  - using the fitted value of $\psi$
- Compare these with observed event times $T_i$ in the treated arm
  - Kaplan-Meier graph
  - Cox model estimates the hazard ratio that would have been observed if the placebo arm was never treated
- Standard error from the Cox model is wrong. Instead
  - use the ITT P-value to construct a test-based CI
  - or bootstrap
Sunitinib overall survival again

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

with thanks to Xin Huang (Pfizer)
Sunitinib overall survival with RPSFTM

- **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

*Estimated by RPSFT model

**Empirical 95% CI obtained using bootstrap samples.

with thanks to Xin Huang (Pfizer)
## Summary: IPCW vs RPSFTM

<table>
<thead>
<tr>
<th></th>
<th>IPCW</th>
<th>RPSFTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assumption</td>
<td>No unmeasured confounders for the decision to switch</td>
<td>Common treatment effect</td>
</tr>
<tr>
<td>Covariate requirements</td>
<td>Anything predicting switch &amp; outcome</td>
<td>None</td>
</tr>
<tr>
<td>Follow-up after switch?</td>
<td>Not needed</td>
<td>Needed</td>
</tr>
<tr>
<td>Handles other treatment changes?</td>
<td>Easily</td>
<td>Need to model &amp; estimate all treatment effects</td>
</tr>
<tr>
<td>Handles censoring?</td>
<td>Easily</td>
<td>Need re-censoring</td>
</tr>
<tr>
<td>Modelling task</td>
<td>Complex (but partly testable)</td>
<td>Simple (but untestable)</td>
</tr>
<tr>
<td>P-value</td>
<td>Often &lt; ITT</td>
<td>Always = ITT</td>
</tr>
</tbody>
</table>
Software

- I wrote **strbee** to fit RPSFTM in Stata
  + forthcoming **strbee2**
- Simon Bond is working on an implementation in R
- Some SAS code snippets are available e.g. Matsuyama (*Stat Med* 2010; 29: 2107–16)
- IPCW is also not hard to code – see Fewell *et al* (cited earlier)
- Miguel Hernan and Jamie Robins have a webpage http://www.hsph.harvard.edu/causal/software/ with
  - STATA and SAS code for fitting marginal structural models \(\rightarrow\) modify for IPCW
  - STATA and SAS code for fitting a nested structural AFT model, which is related to the RPSFTM
Plan

1. Motivation: some recent trials with a big treatment switching problem
2. What’s the question?
3. What methods are available?
4. **Improving the RPSFTM**
5. Implications for practice
Improvement 1: sensitivity analyses

- Can now fit the RPSFTM assuming
  - treatment effect in arm 1 is \( \psi \)
  - treatment effect in arm 0 is \( k\psi \)
  - sensitivity parameter \( k \) is assumed known
- Vary \( k \) over a plausible range, e.g. 0.5 to 1
Improvement 2: more powerful test

- RPSFTM preserves the ITT P-value
- Usually comes from the log rank test
- Can we devise a better (more powerful) test, to be used both in the ITT and RPSFTM analyses?
- Work with Jack Bowden and Shaun Seaman

Recall sunitinib:
P=0.007, 0.107, 0.306 at 1, 2, 4 years.

Power is lost because the treatments received by the arms converge over time.
Weighted log rank test

• Define weighted log rank test statistic for some set of weights $W_j$ for the $j^{th}$ event ($j = 1, \ldots, n$):

$$\sqrt{\sum_j W_j^2 V_j} \sum_j W_j (O_j - E_j)$$

• Reduces to standard test statistic if $W_j = \text{const}$

• The optimal asymptotic choice for weights is $W_j \propto \text{ITT log hazard ratio at time } t_j$ (Schoenfeld, 1981)
  - unweighted test is optimal if hazard ratio is constant

• We derive a simple approximation for $W_j$ (extends method of Lagakos et al, 1990)


Simple approximation for optimal weights

- **Working assumptions:** hazard = $h_{\text{off}}(t)$ whenever off treatment and $h_{\text{on}}(t)$ whenever on treatment
  - $h_{\text{on}}(t) = \theta \ h_{\text{off}}(t)$
  - $\theta \approx 1$
- Let $\gamma^k(t) = P(\text{on treatment at } t \mid T \geq t, Z = k)$
  - recall $Z=\text{arm}, T=\text{time to event}$
- Optimal weight is $W_j \propto \gamma^1(t_j) - \gamma^0(t_j) = \text{difference in proportion of people on treatment in each arm at } j^{\text{th}} \text{ observed event time } t_j$
  - we estimate $\gamma^0(t_j), \gamma^1(t_j)$ and hence $W_j$ from the data
- More theoretical derivation of result exists (Robins, 2011, personal communication)
- Long format $\rightarrow$ weighted log rank test is easy to code
Sunitinib trial: weights and results

- ITT P-values:
  - unweighted $P = 0.31$
  - weighted $P = 0.14$

- RPSFTM analyses:
  - standard $\hat{\psi} = -2.55$  
    $(-3.47, +1.68)$
  - weighted $\hat{\psi} = -0.96$  
    $(-2.47, +0.46)$

- But should negative weights be set to zero?
A small simulation study

<table>
<thead>
<tr>
<th>Setting</th>
<th>Log rank method</th>
<th>ITT</th>
<th>RPSFTM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean $\psi$</td>
<td>p(reject NH)</td>
</tr>
<tr>
<td>$\psi=0$</td>
<td>unweighted</td>
<td>0.000</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>weighted</td>
<td>-0.008</td>
<td>0.04</td>
</tr>
<tr>
<td>$\psi=-0.693$</td>
<td>unweighted</td>
<td>-0.126</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>weighted</td>
<td>-0.435</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Both methods preserve type I error when $\psi=0$
Both methods estimate $\psi$ with small bias
Weighted log rank test is more powerful and more accurate
Plan

1. Motivation: some recent trials with a big treatment switching problem
2. What’s the question?
3. What methods are available?
4. Improving the RPSFTM
5. **Implications for practice**
Key messages: design

- Should trials be designed to avoid switches?
- Clinical perspective: ideal design to evaluate a new drug:
  - drug as it would be used in practice vs.
  - no drug at all
- But
  - ethics often requires drug offered to all
  - this also facilitates recruitment
- Compromise: would it be reasonable to have a 2nd randomisation on progression?
Key messages: data collection

- Collect follow-up data after treatment changes
  - distinguish “withdrawn from treatment” from “withdrawn from the trial”
  - also see US National Research Council report (http://www.nap.edu/catalog/12955.html)
  - needed for ITT and RPSFTM (but not for IPCW)
- Collect covariates that predict whether a patient will switch
  - needed for IPCW & MSM
  - time-dependent covariates
Key messages: trial analysis

- Trialists should pre-specify which method to use (IPCW, RPSFTM, other)
- IPCW: pre-specify
  - definition of switch (at which data will be censored)
  - covariates to be used in modelling switch
  - method for constructing weights
- RPSFTM: pre-specify
  - definition of “on-treatment” variable $D(t)$
  - test to be used
  - re-censoring procedure
- and in both cases, pre-specify baseline covariates to be adjusted for in the analysis (as is done for ITT)
NICE: similar issues with other drugs (1)

- **Pazopanib** for the first-line treatment of advanced renal cell carcinoma
  - analysis used IPCW and RPSFTM
- **Lenalidomide** for the treatment of multiple myeloma in people who have received at least one prior therapy
  - analysis used historical controls
NICE: similar issues with other drugs (2)

- **Trastuzumab + anastrozole** for postmenopausal women with HER2+ and HR+ breast cancer
  - analysis used RPSFTM (2010)
- **Letrozole and anastrozole** vs. tamoxifen as adjuvant therapy in postmenopausal women with early breast cancer
  - analysis used IPCW and RPSFTM (2011)
Summary & questions

• Important to ask the right question
• IPCW and RPSFTM
  – make different assumptions
  – have different strengths
  – have different data requirements
• Best choice of method depends on circumstances
  – can we choose at trial design stage?
  – or must these be post-hoc analyses?
  – simulation studies can’t give definitive answers because the methods make different assumptions
• Consider sensitivity analyses to departures from identifying assumptions
• Can we fix the design?
The end
Further reading

