MRC | Biostatistics Unit

Methods for handling treatment switching: rank-preserving structural nested failure time models, inverse-probability-of-censoring weighting, and marginal structural models

HTMR network workshop on Methods for adjusting for treatment switches in late-stage cancer trials

London, 20th February 2012

Ian White

MRC Biostatistics Unit, Cambridge, UK

Motivation

- Sunitinib trial (from Blay, 2010)
- Big treatment effect on progression-free survival
- Many in placebo arm "switched" to receive sunitinib after progression
- No treatment effect on overall survival (except in very early follow-up)
- How can we analyse such data?
- More later from Xin Huang

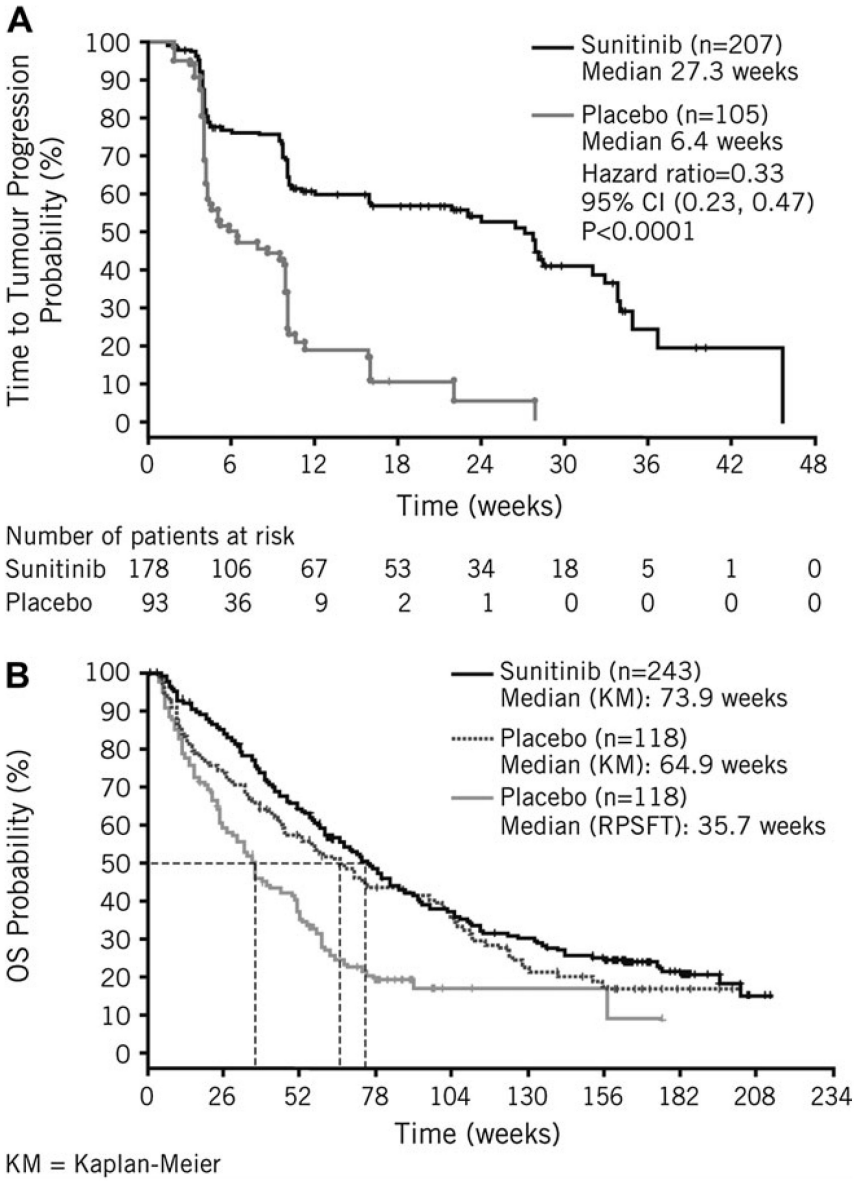


Figure 1. (**A**) Kaplan–Meier estimates of time to tumour progression during the blinded phase of the study and (**B**) Kaplan–Meier and rank-preserving structural failure time (RPSFT) estimates of long-term overall survival (OS) for patients treated with sunitinib or placebo in the phase III study. Results were based on the intent-to-treat population [21, 22].

Note on terminology

- I've called this workshop "Methods for handling treatment switching …"
- Others use "treatment cross-overs" – but may led to confusion with cross-over trials?
- Links to wider statistical literature on "non-compliance"
  - where I'd prefer the non-judgemental "departures from randomised treatment"

Scope of the problem

Many trials have not just treatment switching (i.e. to the treatment allocated to the other trial arm), but also:

- Other changes of prescribed treatment
  - changes to non-trial treatments
  - changes to no treatment
  - multiple treatments
  - dose adjustment
- Non-compliance with prescribed treatment

Plan of the day

Ian White	Methods for handling treatment switching: rank-preserving structural nested failure time models, inverse-probability-of-censoring weighting, and marginal structural models
Susie Dodd	Departure from treatment protocol in published RCTs: a review
James Morden	Methods for adjusting survival estimates in the presence of treatment crossover – simulation studies
Nick Latimer	
Neil Hawkins	Methods for health economic models in metastatic cancer
Xin Huang	Adjusting the Crossover Effect in Survival Analysis Using a Rank Preserving Structural Failure Time Model: The Case of Sunitinib GIST Trial
Rob Hemmings	Treatment switches in cancer trials – problems, pitfalls and (no) solutions
Martin Pitt & Martin Hoyle	Dealing with treatment switches in cost-effectiveness analysis: the NICE experience
Claire Watkins	Discussants & general discussion
Chris Metcalfe	

My talk: introduction to the methods

1. Intention-to-treat analysis
2. Per-protocol analysis
3. Inverse-probability-of-censoring weighting (IPCW)
4. Marginal structural models (MSMs)
5. Rank-preserving structural nested failure time models (RPSFTMs)
6. Brief comparisons

Defining the question

- What is the effect of assignment to treatment A *in the circumstances of the trial*? (effectiveness)
  - could be: A immediately vs. A on progression
- 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
- What is the effect of treatment A *per se* (efficacy)?
  - i.e. while actually given

Defining the question: counterfactuals

- Examples of counterfactual outcomes:
  - the treatment that patient *i* would have had if they had been randomised to treatment A
  - the outcome that would have been observed if patient *i* had received treatment A
- Useful in defining the question: e.g.
  - estimate difference between arms in the subgroup who would take treatment if randomised to it
  - estimate difference between arms if there had been no departures from randomised treatment

A hypothetical (& simplistic) trial

- Randomisation to two arms (Drug A vs placebo)
- Two follow up times
- 1st follow-up detects those whose disease has progressed, but assume no deaths
- Patients in the placebo arm who have progressed are allowed the opportunity to switch to Drug A
- 2nd follow-up looks at mortality (as a %).

- Our question is: what would the difference between the two arms be if no switching occurred?

Hypothetical trial data (observed counts)

Arm	Time 1		Time 2 status
	Progression	Switch	
Drug A	No (800)		Dead (10)
			Alive (790)
	Yes (200)		Dead (90)
			Alive (110)
Placebo	No (600)	No (600)	Dead (10)
			Alive (590)
	Yes (400)	No (200)	Dead (90)
			Alive (110)
		Yes (200)	Dead (30)
			Alive (170)

Intention-To-Treat (ITT) Analysis

- Comparison of outcomes for participants as randomised
  - treatment actually received is ignored in the analysis
- Evaluates the effect of the offer of treatment rather than treatment receipt (so needs fewer assumptions)
  - evaluates effectiveness as opposed to efficacy

- Essential part of analysis
  - an unbiased answer
  - but possibly to the wrong question
- At least, we'd need to know amounts of treatments actually received to interpret the results of ITT analysis
  - topic of Susie Dodd's talk

Hypothetical trial: ITT analysis

Arm	Time 1		Time 2 status
	Progression	Switch	
Drug A	No (800)		Dead (10)
			Alive (790)
	Yes (200)		Dead (90)
			Alive (110)
Placebo	No (600)	No (600)	Dead (10)
			Alive (590)
	Yes (400)	No (200)	Dead (90)
			Alive (110)
		Yes (200)	Dead (30)
			Alive (170)

ITT: 100/1000 vs 130/1000

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: PP analysis

Arm	Time 1		Time 2 status
	Progression	Switch	
Drug A	No (800)		Dead (10)
			Alive (790)
	Yes (200)		Dead (90)
			Alive (110)
Placebo	No (600)	No (600)	Dead (10)
			Alive (590)
	Yes (400)	No (200)	Dead (90)
			Alive (110)
		Yes (200)	Dead (30)
			Alive (170)

PP: 100/1000 vs 100/700

Inverse-probability-of-censoring weighting (IPCW) methods

- Robins & Finkelstein (2000)
- Like per-protocol analysis, IPCW views outcome data collected after a treatment switch as irrelevant
- Follow-up data (time of death, for instance) are artificially censored at the time of treatment switch
- A model is constructed to predict this artificial censoring (= treatment switching)
  - must include all baseline or post-randomisation variables that both predict treatment switching and outcome: "no unmeasured confounders"
  - hard to be confident that we have done this
- NB two models:
  - "switching model" to predict switching
  - main interest is in "outcome model": e.g. Cox model for death on randomised group

IPCW analysis showing **weights**

Arm	Time 1		Time 2 status
	Progression	**Switch**	
Drug A	No (800)		Dead (10) **1**
			Alive (790) **1**
	Yes (200)		Dead (90) **1**
			Alive (110) **1**
Placebo	No (600)	No (600)	Dead (10) **1**
			Alive (590) **1**
	Yes (400)	No (200)	Dead (90) **2**
			Alive (110) **2**
		Yes (200)	Dead (30) **0**
			Alive (170) **0**

IPCW: 100/1000 vs 190/1000

Inverse probability weights

- Half of our progressing placebo patients switched to drug A
- The other half did not switch
- Assume the switch-free outcome in both of these two groups is similar (independent of switching)
  - here I'm talking about a counterfactual outcome
- Then we can use the non-switchers' data but weight it by a factor of 2 to represent the switchers' data *if they had not switched*

Constructing inverse probability weights

- Need a model for artificial censoring (=switching) given baseline and time-dependent covariates
- Switching models:
  - discrete time: logit P(switch at time t) = $\alpha_t + \beta' \mathbf{X_t}$
  - continuous time: $h(t) = h_0(t) \exp(\beta' \mathbf{X_t})$
- Fit switching model & hence estimate p_{it} = P(individual i has not yet switched by time t) for all outcome-event times t
- Weight the analysis of the outcome model by $w_{it} = 1/p_{it}$
  - time-dependent weights are a problem in some software
  - need robust (sandwich) standard errors to allow for the weighting

Choice of covariates for IPCW

- Recall: anything that predicts both switching and outcome
- Baseline covariates: the usual stuff?
- Time-dependent covariates:
  - progression
  - severity (performance status etc.)
  - anything you think clinicians would use to decide whether to switch (need to speak with a clinician)
- Time-dependent covariates are very important

The problem of unstable weights

- Sometimes we get very large weights in IPCW
  - e.g. if 99% of patients who progressed then switched, the poor 1% who didn't switch get a weight of 100 to "represent" those who did switch
- Leads to large standard errors (small effective sample size)
- "Capping" weights avoids large standard errors but re-introduces bias
- "Stabilised" weights can help (Robins et al, 2000)
- Inherent limitation of the method
  - e.g. if 100% of patients who progressed then switched, IPCW simply fails

Note on IPCW

- Can handle more than just switching – e.g.
  - IPCW applies for any sort of treatment changes
  - can also use it for other "protocol violations" such as loss to follow up
- Core assumption must be re-assessed for each new application

IPCW summary

1. Identify important baseline and time-dependent covariates that predict both switching and outcome
2. Model the probability of switching at each time given covariates
3. For each individual and each time, calculate their probability of remaining unswitched given their covariates
4. For the unswitched, calculate time-dependent weights as the inverse probability of remaining unswitched
   - optionally stabilised weights
5. Fit a Cox model of survival on randomised group with time-dependent weights to the data, censoring at time of switch

Marginal Structural Models (MSMs)

- Similar idea to IPCW (Robins et al, 2000)
- IPCW compares two potential treatment histories:
  - treated at start (identified from the treatment arm)
  - never treated (identified from the weighted placebo arm, censored at treatment)
- MSMs compare a wider range of potential treatment histories, e.g.
  - treated from progression
  - treated for *d* months
- MSM is a model for causal effects across potential treatment histories
  - e.g. causal effect of treatment for *d* months = βd
- The model is estimated by weighting the data to estimate outcomes under each potential treatment history

Causal effect: comparison of counterfactual outcomes (given different potential treatment histories) in the **same** individuals

MSM analysis showing **weights**

Arm	Time 1		Time 2 status	
	Progression	Switch		
Drug A	No (800)		Dead (10)	1
			Alive (790)	1
	Yes (200)		Dead (90)	1
			Alive (110)	1
Placebo	No (600)	No (600)	Dead (10)	1 1
			Alive (590)	1 1
	Yes (400)	No (200)	Dead (90)	2 0
			Alive (110)	2 0
		Yes (200)	Dead (30)	0 2
			Alive (170)	0 2

Treated from start: 100/1000

Never treated: 190/1000

Treated from progression: 70/1000

Towards the RPSFTM

- focussing on time-to-event 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
    - » *may be time-dependent*
  - T_i: observed outcome (time to event)
- Ignore censoring for now
- *Counterfactual* or *potential* outcome $T_i(0)$
  - outcome that *would have been observed* without treatment
- The RPSFTM relates T_i to $T_i(0)$ through a treatment effect ψ (psi)

RPSFTM (2)

- The RPSFTM relates observed outcome T_i to treatment-free outcome $T_i(0)$ through a treatment effect ψ

- Case 1: all-or-nothing treatment (e.g. surgical intervention)
  - untreated individuals: $T_i = T_i(0)$
  - treated individuals: $T_i = \exp(-\psi) \times T_i(0)$
  - or $T_i(0) = \exp(\psi) \times T_i$
  - treatment multiplies lifetime by a ratio $\exp(-\psi)$
  - $\psi<0$ means treatment is good

"Rank-preserving": if i dies before j when both are treated, then i dies before j when both are untreated

RPSFTM (3)

- Case 2: time-dependent 0/1 treatment (e.g. drug prescription, ignoring actual adherence)
  - Define T_i^{off}, T_i^{on} as times off and on treatment
    - » *so* $T_i^{off} + T_i^{on} = T_i$
  - Treatment multiplies just the T_i^{on} part of the lifetime
    - » *time* T_i^{on} *on treatment "equals" time* $\exp(\psi) \times T_i^{on}$ *off treatment*
  - Model: $T_i(0) = T_i^{off} + \exp(\psi) \times T_i^{on}$

- Case 3: time-dependent quantitative treatment (e.g. drug adherence)
  - can still define ALM, but it's more complicated:

$$T_i(0) = \int_0^{T_i} \exp\{\psi D_i(t)\} dt$$

Interpretation of ψ

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

- Treatment multiplies lifetime by a ratio $\exp(-\psi)$
- Best interpreted in terms of an ageing or disease process: e.g. tumour is growing but drug doubles the time it takes to grow a given amount [if $\exp(\psi)=0.5$]
- $\exp(\psi)$ sometimes called an acceleration factor – factor by which your life is speeded up – or a time ratio

- But I'll show later that you don't have to interpret ψ

RPSFT model assumptions

- Common treatment effect
  - treatment effect, expressed as ψ, is the same for control arm (treated from progression) as for experimental arm (treated from randomisation)
- Exclusion restriction
  - untreated outcome T(0) is independent of randomised group Z
- Comparability of switchers & non-switchers is NOT assumed

G-estimation: an unusual estimation procedure

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

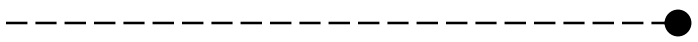
- Take a range of possible values of ψ
- For each value of ψ, work out T(0) and test whether it is balanced across randomised groups
- Graph test statistic against ψ
- Best estimate of ψ is where you get best balance (smallest test statistic)
- 95% CI is values of ψ where test doesn't reject

- You can choose which test to use!
- Conventionally the same test as in the ITT analysis
  - usually log rank test or adjusted Cox model
  - we're researching possible power gains from other choices

Illustration of the model

Suppose $e^{\psi} = 0.5$ – so 1 year on treatment "equals" 0.5 years off treatment.

Possible outcomes for a subject with T(0) = 1 year:

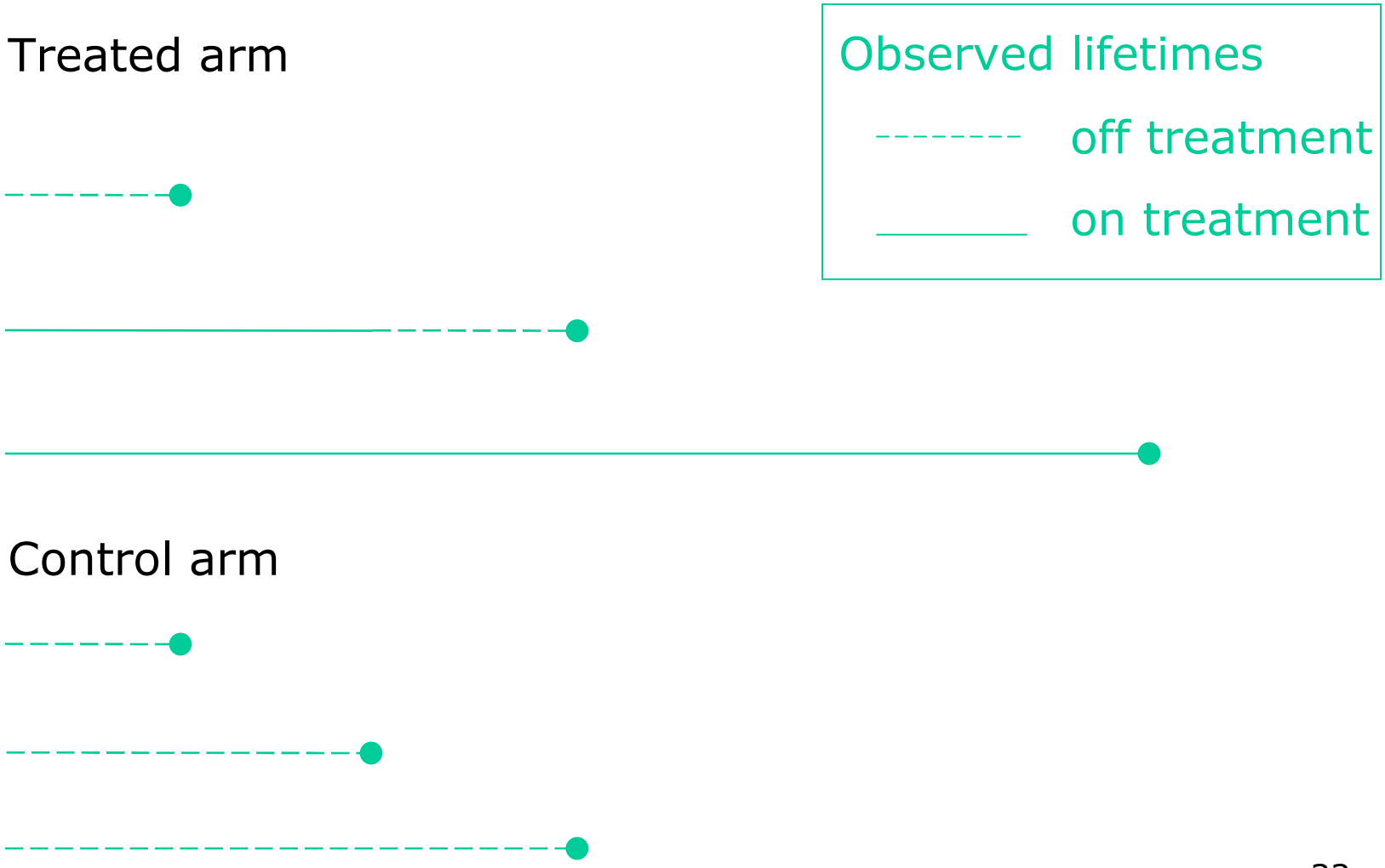


If completely untreated, life = 1 year

If treated for 1 year, life = 1.5 years

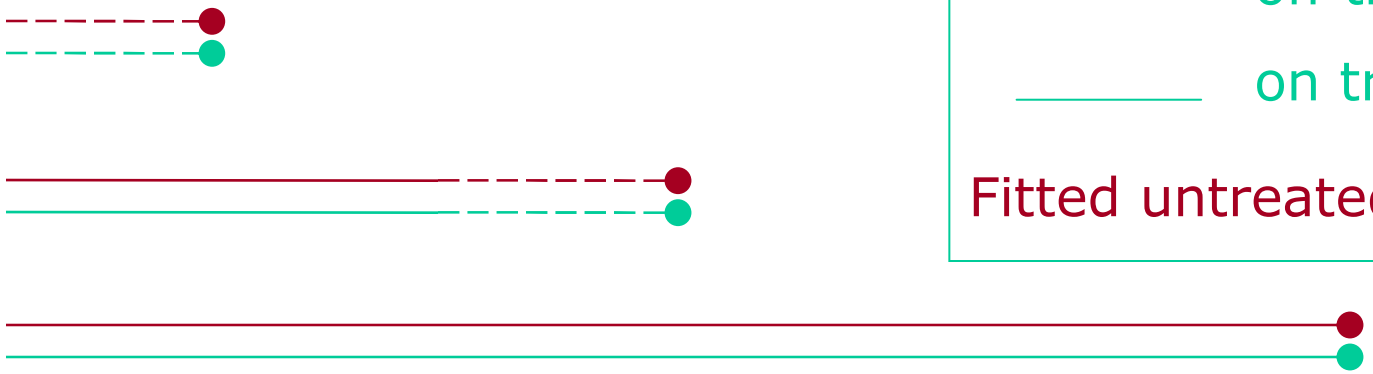
If completely treated, life = 2 years

Hypothetical data (switches occur only in treated arm)



Estimating ψ: is $e^{\psi} = 1$?

Treated arm



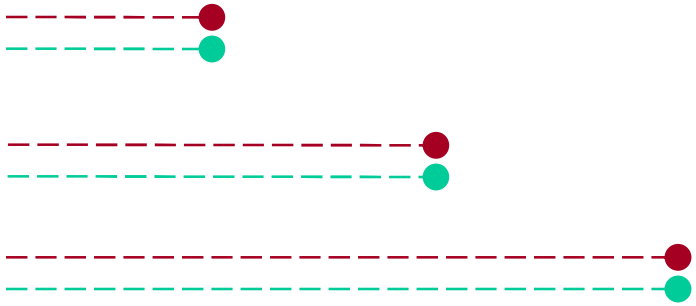
Observed lifetimes

off treatment

on treatment

Fitted untreated lifetime

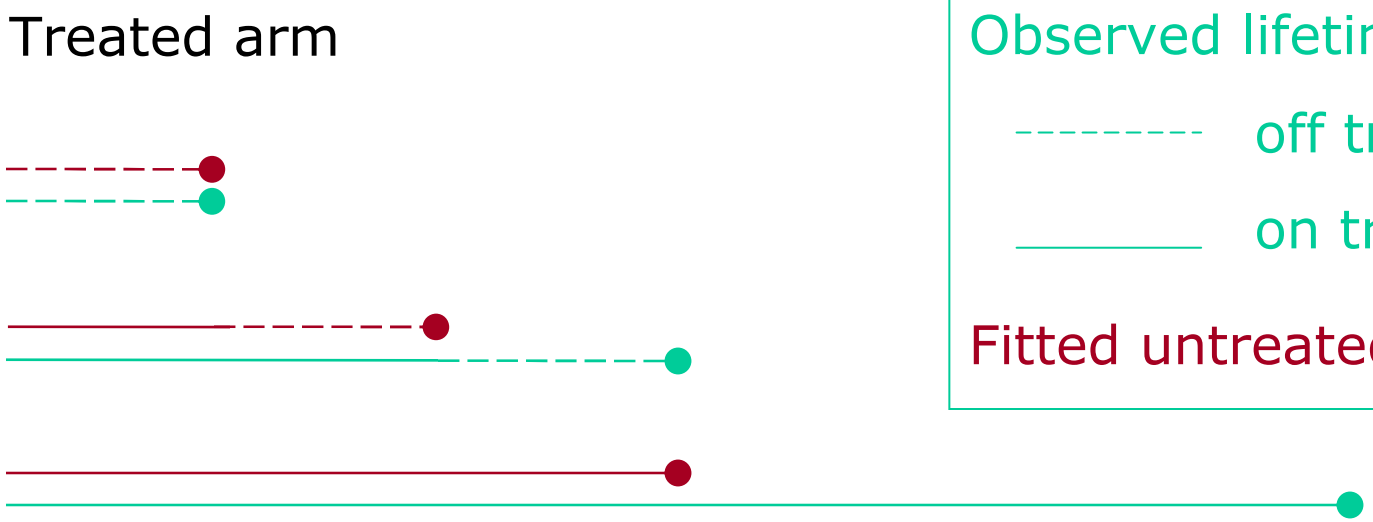
Control arm

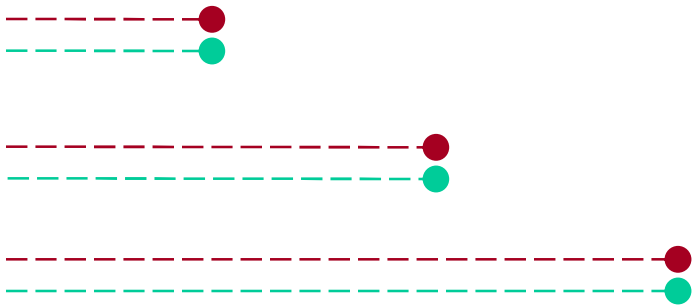


If $e^{\psi} = 1$ *then* untreated lifetimes differ between arms

So estimated $e^{\psi} \neq 1$

Estimating ψ: is $e^{\psi} = 0.5$?





If $e^{\psi} = 0.5$ *then*
untreated lifetimes
balance between arms

So estimated $e^{\psi} = 0.5$

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
- The estimation procedure is "randomisation-respecting"
  - it is based only on the comparison of groups as randomised
  - I've suggested the term randomisation-based efficacy estimator, RBEE (White, 2005)

Censoring

- Censoring introduces complications in RPSFTM estimation
  - censoring on the T(0) scale is informative
  - requires *re-censoring* which can lead to strange results – see White *et al* (1999)

Estimating a causal hazard ratio

- Often hard to interpret ψ
- Use the RPSFTM again to estimate the untreated event times $T_i(0)$ in the placebo arm
  - using the fitted value of ψ
- Compare these with observed event times T_i in the treated arm
- Use a Kaplan-Meier graph and Cox model
- Cox model estimates the hazard ratio that would have been observed if the placebo arm was never treated
- **Don't** use the P-value / CI from the Cox model – it is much too small. Instead
  - use the ITT P-value to construct a test-based CI
  - or bootstrap (White et al, 1999)

Non-standard RPSFTM-like analysis to estimate a relative treatment effect θ

Arm	Time 1		Time 2 status	Deaths if untreated
	Progression	Switch		
Drug A	No (800)		Dead (10)	$10/\theta$
			Alive (790)	
	Yes (200)		Dead (90)	$90/\theta$
			Alive (110)	
Placebo	No (600)	No (600)	Dead (10)	10
			Alive (590)	
	Yes (400)	No (200)	Dead (90)	90
			Alive (110)	
		Yes (200)	Dead (30)	$30/\theta$
			Alive (170)	

Assumes only current treatment matters

Solve $10/\theta + 90/\theta = 10 + 90 + 30/\theta \Rightarrow \theta = 0.70$

Summary: IPCW vs RPSFTM

	IPCW	RPSFTM
Assumption	No unmeasured confounders	Common treatment effect
Covariate requirements	Anything predicting switch & outcome	None
Follow-up after switch?	Not needed	Needed
Handles other treatment changes?	Easily	Difficult
Modelling task	Complex (but partly testable)	Simple (but untestable)
Power	Often > ITT	Same as ITT

To be compared in talk by James Morden & Nick Latimer

Arguments used: IPCW

For

- Gives HR rather than acceleration factor
- Does not borrow information from switched patients
- More powerful than ITT
- Does not model the effect of cross-over

Against

- Assumes no unmeasured confounders for the decision to switch
- Do we understand why some patients do not switch after progression?

Arguments used: RPSFTM

For

- Preserves ITT P-value
- Don’t need no-unmeasured-confounders assumption
- Valid under non-ignorable (selective) selection to switch
- No need to model covariate effects

Against

- Need to model all treatment effects – awkward for comparative trials where treatments may stop, and for trials with second-line treatments
- Re-censoring

Key messages: design

- Collect follow-up data after treatment changes
  - distinguish "withdrawn from treatment" from "withdrawn from the trial"
  - also see new 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 cross over
  - needed for IPCW & MSM
  - time-dependent covariates

Key messages: pre-specification

- Pre-specify which method to use (IPCW, RPSFTM, other)
- IPCW: pre-specify
  - definition of cross-over (at which you will censor)
  - covariates to be used in modelling cross-over
  - 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 you do for ITT)

Summary & questions

- IPCW, MSM and RPSFTM
  - make different assumptions
  - have different strengths
  - have different data requirements
- Best choice depends on circumstances
- How should we choose?
  - can we do so at trial design stage?
  - or must these be post-hoc analyses?
- Should we consider sensitivity analyses?
- Can we improve on these methods? (hybrids??)

References

- Blay JY. Pharmacological management of gastrointestinal stromal tumours: an update on the role of sunitinib. *Annals of Oncology* 2010; **21**: 208–215.
- Greenland S, Lanes S, Jara M. Estimating effects from randomized trials with discontinuations: the need for intent-to-treat design and G-estimation. Clinical Trials 2008; **5**: 5. *Tutorial on RPSFTM claiming it should become a standard procedure.*
- Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000; **11**: 561–570. *MSMs for Cox model.*
- 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. *Introduced IPCW.*
- Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; **11**: 550–560. *Introduced MSMs & stabilised weights.*
- Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics – Theory and Methods 1991; **20**: 2609–2631. *Introduced RPSFTM.*
- White IR. Uses and limitations of randomization-based efficacy estimators. Statistical Methods in Medical Research 2005; **14**: 327–347. *Review.*
- White IR, Babiker AG, Walker S, Darbyshire JH. Randomisation-based methods for correcting for treatment changes: examples from the Concorde trial. Statistics in Medicine 1999; **18**: 2617–2634. *Exploration of RPSFTM.*