The statistics of cancer survival - What are the true survival benefits associated with new cancer treatments?

Nicholas Latimer, University of Sheffield
Collaborators: Paul Lambert, Keith Abrams, Michael Crowther, Allan Wailoo and James Morden
Contents

- Health Economics
- What is treatment crossover and why does it occur?
- What problems are created by treatment crossover?
- How important is treatment crossover?
- What are the potential solutions?
- Simulation study
- Conclusions
Health Economics

- Appraise new treatments to see if they are “cost-effective”

⇒ Should the NHS buy them?

- NHS has a fixed budget – has to try to maximise health benefits by buying the most cost-effective treatments

⇒ Need a generic outcome measure – the QALY

⇒ And a cost-effectiveness threshold

- Need to estimate costs and QALYs accurately so that consistent decisions can be made

⇒ For cancer treatments, survival is likely to be key
Treatment crossover (1)

- In RCTs often patients are allowed to switch from the control treatment to the new intervention after a certain timepoint (e.g., disease progression)
  - PFS (progression free survival) estimates are ok
  - But OS (overall survival) estimates will be confounded

What are the implications of this?

- For clinical analysis
- For economic analysis
  - There are different drivers for these two analyses
Treatment crossover (2)

- **Clinical analysis**
  - Drug regulatory bodies such as FDA and EMA accept that PFS is sufficient for licensing
  - There are reduced incentives for companies to collect longer term survival data
  - There are reduced incentives to maintain randomisation post-progression
  - ➔ Practical reason why treatment crossover occurs
  - ➔ Combined with ethical reasons, strong incentives to allow crossover

- **Economic analysis**
  - For interventions that impact upon survival OS is a key input in the economic model
  - Need accurate estimates of the treatment effect on PFS and OS
Treatment crossover (3)

- Treatment crossover is **not** just an issue for economic evaluation.

- But it can appear that way because it becomes more of an issue at the “fourth hurdle”.

**Implications:**

- Cost effectiveness results will be inaccurate → an ITT analysis is likely to underestimate the treatment benefit.

- Inconsistent and inappropriate treatment recommendations could be made.
Treatment crossover (4)

Crossover is likely to result in an underestimate of the treatment effect.
What is usually done to adjust?

No clear consensus

Numerous ‘naive’ approaches have been taken in NICE appraisals:
- Take no action at all
- Exclude or censor all patients who crossover

Occasionally more complex statistical methods have been used, eg:
- Rank Preserving Structural Failure Time Models (RPSFTM)
- Inverse Probability of Censoring Weights (IPCW)

And others are available from the literature, eg:
- Structural Nested Models (SNM)
What are the consequences?

NICE TA 215, Pazopanib for RCC [51% of control switched]

- **ITT:** OS HR (vs IFN) = 1.26 \(\rightarrow\) ICER = Dominated
- **Censor patients:** HR = 0.80 \(\rightarrow\) ICER = £71,648
- **Exclude patients:** HR = 0.48 \(\rightarrow\) ICER = £26,293
- **IPCW:** HR = 0.80 \(\rightarrow\) ICER = £72,274
- **RPSFTM:** HR = 0.63 \(\rightarrow\) ICER = £38,925
Potential solutions (1)

**RPSFTM**
- Developed for use on RCT datasets, makes use of randomisation to estimate counterfactual survival times

**Key assumption:** common treatment effect

**IPCW**
- Developed for use on observational datasets, censors xo patients, weights remaining patients, runs weighted Cox model

**Key assumptions:** “no unmeasured confounders”; must model OS and crossover using covariate data

**SNM**
- Observational version of RPSFTM

**Key assumptions:** “no unmeasured confounders”; must model OS and crossover
Potential solutions (2)

Another option…

- Consider the treatment crossover typically seen in oncology trials…
- Data on PFS is required for licensing, thus only allow crossover post-progression
- If crossover only happens after progression, and happens soon after progression, we may consider a simple “two-stage” approach:
  - Use disease progression as a secondary baseline for control group patients and consider control group data after this time-point as an observational dataset
  - Apply an accelerated failure time model to this dataset including covariates for crossover and other prognostic covariates measured at the secondary baseline
  - Use the AF derived for crossover to “shrink” survival times of switchers
  - Counterfactual dataset

Key assumptions: “no unmeasured confounders” at secondary baseline time-point; crossover only after progression, and soon after progression
Simulation study (1)

- None of these methods are perfect
- But we need to know which are likely to produce least bias in different scenarios

Simulation study

- Simulate survival data for two treatment groups, applying crossover that is linked to patient characteristics/prognosis
- In some scenarios simulate a treatment effect that changes over time
- In some scenarios simulate a treatment effect that remains constant over time
- Test different %s of crossover, and different treatment effect sizes

How does the bias and coverage associated with each method compare?
Simulation study (2)

Methods assessed

- **Naive methods**
  - ITT
  - Exclude crossover patients (PPexc)
  - Censor crossover patients (PPcens)
  - Treatment as a time-dependent covariate (TDCM)

- **Complex methods**
  - RPSFTM
  - IPE algorithm
  - IPCW
  - SNM
  - Two-stage Weibull
Results: common effect

- RPSFTM / IPE worked very well
- IPCW and SNM performed ok when crossover % was lower
- IPCW and SNM performed poorly when crossover % was very high
- Naive methods performed poorly (generally led to higher bias than ITT)
- Two-stage Weibull performed well
Results: effect 15% ↓ in xo patients

- RPSFTM / IPE produced higher bias than previous scenarios
- IPCW and SNM performed similarly to RPSFTM / IPE providing crossover < 90%
- IPCW and SNM performed poorly when crossover % was very high
- Bias not always lower than that associated with the ITT analysis
- Two-stage Weibull performed well
Results: effect 25% ↓ in xo patients

- RPSFTM / IPE produce substantial bias
- IPCW and SNM produce less bias than RPSFTM / IPE providing crossover < 90%
- Few ‘good’ options when crossover % is very high
- Often ITT analysis likely to result in least bias (esp. when trt effect low)
- But two-stage Weibull still does quite well
Results cont.

Relationship between bias and treatment crossover %

-40
-20
0
20
40
60
80
100
120
140

Mean % bias

Crossover proportion (at-risk patients)

ITT
Results cont.

Relationship between bias and treatment crossover %
Results cont.

Relationship between bias and treatment crossover %

Mean % bias vs. Crossover proportion (at-risk patients)

- IPCW
- IPE
- RPSFTM
- ITT
Results cont.

Relationship between bias and treatment crossover %

Mean % bias vs. Crossover proportion (at-risk patients) for different methods:
- IPCW
- IPE
- RPSFTM
- SNM
- ITT
Results cont.

Relationship between bias and treatment crossover %
How can we select the most appropriate method?

- Even the more complex methods have important limitations and will often result in bias in realistic scenarios.
- “Naive” methods should not be used.

1. What was the crossover mechanism? Who, when, why and how many?
2. What is the nature of the treatment effect?
3. What / how much data are available? Important time-dependent covariates?

- Each of these questions helps determine whether ITT, RPSFTM, IPE, IPCW or two-stage methods are likely to be suitable.
- How plausible are their assumptions in an oncology RCT context?
Limitations

- Other scenarios would be interesting
  - Lower crossover proportions
  - Different sample sizes
  - Different treatment effect decrements

- Data generating model
  - We used a joint longitudinal and survival model starting off with a Weibull distribution
  - Does this influence the results?

- New methods are required!
Conclusions (1)

- Treatment crossover is an important issue that has come to the fore in HE arena

- Current methods for dealing with treatment crossover are imperfect and have been used uncertainly in HTA

- Our study offers evidence on bias in different scenarios (subject to limitations)
Conclusions (2)

- RPSFTM / IPE produce low bias when treatment effect is common
  - But are very sensitive to this

- IPCW / SNM are not affected by changes in treatment effect between groups, but in (relatively) small trial datasets observational methods are volatile
  - Especially when crossover % is very high (leaving low \( n \) in control group)

- Simple two-stage methods are worthy of consideration

- Very important to assess trial data, crossover mechanism, treatment effect to determine which method likely to be most appropriate
  - There is a definite requirement for clinical opinion, to determine justifiable methods

- Don’t just pick one!!