Statistical, clinical and ethical considerations when minimizing confounding for overall survival in cancer immunotherapy trials

Dominik Heinzmann, PhD
Global Development Team Leader HER2
Associate Director Biostatistics, Roche, Basel
## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment switch – What is the estimand?</td>
</tr>
<tr>
<td>Challenges in CIT endpoints and treatment switch implications</td>
</tr>
<tr>
<td>Outlook</td>
</tr>
</tbody>
</table>
Table of Contents

- Treatment switch – What is the estimand?
- Challenges in CIT endpoints and treatment switch implications
- Outlook
Non-Proportional Hazards (NPH): What does It mean?

• Most popular methods for analysis of time to event trials:
  – log-rank test (testing) – Proof of efficacy
  – Cox regression (estimation) – Quantify treatment effect

• Hazard ratio, naive median differences and milestone survival differences are standard way of summarizing treatment effect

• Are they appropriate summary measures when the treatment effect is not constant over time (eg NPH situation)?
Non-Proportional Hazards (NPH): What does It mean?

• Different types of NPH
  1. Delayed treatment effect
  2. Diminishing treatment effect (eg treatment switch control arm)
  3. Crossing Hazard
  4. Long term survivor (“cure” / long-term survival rate)
  5. Subgroup effect: NPH driven by particular subgroup

• Combination of different types is possible (e.g., Type 4 can occur in combination with 1)
Treatment switch situation with CIT

- For many CIT trials, patients can only be enrolled if they did not obtain previous CIT.
- Hence only control arm patients from e.g. a first line CIT trial could participate in subsequent line CIT trials, impacting on the likelihood to capture the benefit of CIT (if it exists).
- For blinded trials (mainly combination trials), this issue may become an ethical dilemma, as keeping the blinding prevents control arm patients to access experimental clinical CIT studies.
Illustrative example: HERA trial (non CIT)

Median follow-up (% follow-up time after selective crossover)

<table>
<thead>
<tr>
<th>Year</th>
<th>Follow-up Time</th>
<th>DFS Benefit</th>
<th>No. of DFS events</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>1 yr MFU (0%)</td>
<td>0.54</td>
<td>127 vs 220, P&lt;0.0001</td>
</tr>
<tr>
<td>2006</td>
<td>2 yrs MFU (4.3%)</td>
<td>0.64</td>
<td>218 vs 321, P&lt;0.0001</td>
</tr>
<tr>
<td>2008</td>
<td>4 yrs MFU (33.8%)</td>
<td>0.76</td>
<td>369 vs 458, P&lt;0.0001</td>
</tr>
<tr>
<td>2012</td>
<td>8 yrs MFU (48.6%)</td>
<td>0.76</td>
<td>471 vs 570, P&lt;0.0001</td>
</tr>
<tr>
<td>2015</td>
<td>11 yrs MFU (=50%)</td>
<td>0.76</td>
<td>505 vs 608, P&lt;0.0001</td>
</tr>
</tbody>
</table>

DFS benefit

HR (95% CI)

Favours 1 year trastuzumab

Favours observation

No. of DFS events
1 year trastuzumab vs observation
Adjust for switch?

1) **[Naïve] Censoring patients** at time of switch (biased decision to switch is usually not independent of prognosis)

2) **Inverse Probability Censoring Weighting (IPCW)**
   a) Creates a scenario of **missing follow-up data** by censoring the follow-up of each patient at the time of crossover
   b) BUT patients are weighted according to their probability to cross-over
   c) A patient will be assigned a weight of > 1 if other patients with similar characteristics crossed over to “re-create” the population that would have been observed without crossover
   d) Weights are based on factors affecting a patient’s decision to cross over or prognostic of survival

**Challenge:** Assumes no unmeasured confounders (i.e. everything predicting switch / OS is collected)

3) **Rank-preserving structural failure time model (RPSFT)**
   a) It works by “re-creating” the survival time of patients, as if they had never received experimental treatment, i.e. patient who switches treatment has a **counterfactual event time** – the time-to-event if no experimental treatment had been received
Illustrative example: HERA trial (non CIT)

Median follow-up (% follow-up time after selective crossover)

- 2005 (0%) 1 yr MFU
- 2006 (4.3%) 2 yrs MFU
- 2008 (33.8%) 4 yrs MFU
- 2012 (48.6%) 8 yrs MFU
- 2015 (=50%) 11 yrs MFU

DFS benefit

- 0.54
- 0.64
- 0.76

No. of DFS events 1 year trastuzumab vs observation

- 2005: 127 vs 220, P<0.0001
- 2006: 218 vs 321, P<0.0001
- 2008: 369 vs 458, P<0.0001
- 2012: 471 vs 570, P<0.0001
- 2015: 505 vs 608, P<0.0001

“Best guess of Treatment effect”?

RPSFT adjustment 0.62 (0.50, 0.76)
Wait a minute - What is the estimand?

- **Intercurrent events**: Causing missing information as to the situation when patients would have adhered to randomized treatment and to assessment as per protocol until end of trial

- Change in protocol treatment: *Treatment switching*

- **Hypothetical estimand**: Effect “when no control patient would have switched to experimental treatment”

- For OS, always subsequent therapies
  - Hence “non-adherence” really intercurrent events?

- In such cases intervention effect: *Treatment policy estimand*
Wait a minute - What is the estimand?

- Current CIT landscape, deal with experimental CIT treatments in subsequent lines
- So back to: Hypothetical estimand?
- OR avoid switch (keep blinding to avoid entering later line CIT trials)
  - “Clear” estimand?

- Remark: Complicated if we have different CIT approvals in later lines in different regions (eg US vs EU)
  - “Clear”/hypo estimand relevant for one region, treatment policy estimand for other region(s)
Table of Contents

- Treatment switch – What is the estimand?
- Challenges in CIT endpoints and treatment switch implications
- Outlook
## Effect magnitude of endpoints with CITs

<table>
<thead>
<tr>
<th>Study name</th>
<th>Indication</th>
<th>Control</th>
<th>Experimental</th>
<th>HR PFS</th>
<th>HR OS</th>
<th>ORR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Checkmate 017</td>
<td>NSCLC 2L</td>
<td>Docetaxel</td>
<td>Nivolumab</td>
<td>0.63 (0.48,0.83)</td>
<td>0.62 (0.48,0.81)</td>
<td>9% vs 20%</td>
</tr>
<tr>
<td>Checkmate 057</td>
<td>NSCLC 2L non-sq</td>
<td>Docetaxel</td>
<td>Nivolumab</td>
<td><strong>0.92</strong> (0.77,1.11)</td>
<td><strong>0.72</strong> (0.60,0.88)</td>
<td>19% vs 12%</td>
</tr>
<tr>
<td>CHECKMATE 026</td>
<td>NSCLC 1L PDL1+</td>
<td>Chemo</td>
<td>Nivolumab</td>
<td>1.15 (0.91,1.45)</td>
<td>1.02 (0.80,1.30)</td>
<td></td>
</tr>
<tr>
<td>KEYNOTE 010</td>
<td>NSCLC 2L</td>
<td>Docetaxel</td>
<td>Pembrolizumab</td>
<td><strong>0.88</strong> (low dose) 0.79 (high dose)</td>
<td><strong>0.71</strong> 0.61</td>
<td>9% vs 18% 9% vs 18%</td>
</tr>
<tr>
<td>KEYNOTE 024</td>
<td>NSCLC 1L PDL1+</td>
<td>Platinum-containing chemo</td>
<td>Pembrolizumab</td>
<td>0.50 (0.37,0.68)</td>
<td>0.60 (0.41,0.89)</td>
<td>28% vs 45%</td>
</tr>
<tr>
<td>POPLAR</td>
<td>NSCLC 2L</td>
<td>Docetaxel</td>
<td>Atezolizumab</td>
<td><strong>0.94</strong></td>
<td><strong>0.73</strong></td>
<td>12% vs 19%</td>
</tr>
<tr>
<td>OAK</td>
<td>NSCLC 1L</td>
<td>Docetaxel</td>
<td>Atezolizumab</td>
<td><strong>0.95</strong> (0.62,0.87)</td>
<td><strong>0.73</strong> (0.62,0.87)</td>
<td>13% vs 14%</td>
</tr>
</tbody>
</table>

**Remark:** Suggestive that PFS may not be the optimal (most sensitive) endpoint?
OS only reliable EP? - Trial implications

- For many CIT trials, patients can only be enrolled if they did not obtain previous CIT.
- Hence only control arm patients from e.g. a first line CIT trial could participate in subsequent line CIT trials, impacting on the likelihood to capture the benefit of CIT (if it exists).
- For blinded trials (mainly combination trials), this issue can become an ethical dilemma, as keeping the blinding prevents control arm patients to access experimental clinical CIT studies.
## OS only reliable EP? - Trial implications

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Percentage of switchers (after PFS)</th>
<th>Resulting diminished power</th>
</tr>
</thead>
<tbody>
<tr>
<td>No switch</td>
<td>0%</td>
<td>-</td>
</tr>
<tr>
<td>Low Switch</td>
<td>30%</td>
<td>-16%</td>
</tr>
<tr>
<td>Medium Switch</td>
<td>50%</td>
<td>-33%</td>
</tr>
<tr>
<td>High Switch</td>
<td>80%</td>
<td>-58%</td>
</tr>
</tbody>
</table>

### Summary impact on OS

- OS results likely heavily impacted by treatment switch of control arm patients to experimental next-line CIT trials (significant power decrease by 16-58%)
- Study would be underpowered for co-primary endpoint OS
OS only reliable EP? - Trial implications

Prohibiting unblinding -> “Clear” estimand
1. Affects the ability of patients/physicians to determine eligibility for participation in subsequent-line experimental CIT trials
2. Does not affect options for treatment with standard-of-care agents
3. Preserves the ability of the CIT study to detect OS and hence provide a new treatment option

Permitting unblinding -> Hypothetical or treatment policy estimand?
1. Compromises ability of the CIT study to detect OS (due to imbalances in subsequent line-therapies between the two arms), potentially negating a future treatment option
2. Maximizes the treatment options (including experimental treatment options) available for each individual patient

Other CIT specific consideration…
1. Treatment beyond progression (“pseudo-progression”)
   • Unethical to treat blinded (placebo & chemo) after RECIST progression
   • Pseudo-progression not entirely objective assessment, cannot be fully controlled (eg may be misused to determine eligibility for subsequent CIT trials)
Interactions with health authorities

Questions:
- (1) blinded? -> “Clear” estimand
- (2) if not adjust model-based adjustment for treatment switch -> Hypothetical estimand
  - Inverse Probability Censoring Weighting (IPCW)
  - Rank-preserving structural failure time model (RPSFT)

General feedback:
- “… but final decision remains with sponsor…”
- Neither FDA nor EMA agreed to use any currently available methods to “adjust” for treatment switch as primary analysis for OS -> not hypothetical estimand
Interactions with external ethics consultants

• **General comment:** Clinical trials are conducted in general to investigate *experimental treatments*. Enrolled patients do no have a guaranteed benefit for themselves, but potentially help future patients.

• **Trial shall be blinded,** but that the *Informed Consent Form* shall clearly state that participation in the trial may prohibit patients to join experimental CIT trials in subsequent lines.

• **Rational:** Ethical requirements that
  
  – clinical research must lead to improvements in health or advancements in generalizable knowledge
  
  – clinical research must produce reliable and valid data that can be interpreted.
  
  – Invalid research includes underpowered studies and studies with biased endpoints
<table>
<thead>
<tr>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment switch – What is the estimand?</td>
</tr>
<tr>
<td>Challenges in CIT endpoints and treatment switch implications</td>
</tr>
<tr>
<td>Outlook</td>
</tr>
</tbody>
</table>
Outlook

• One may observe more and more approvals of CITs in a shorter time frame (eg through breakthrough and AA in US, PRIME and CMA in EU…)

• What if different CITs (eg PD-1 after PD-L1…) are used across lines, is there a hope to still being able to measure benefit in a trial?
  – PFS for combo sensitive? We need to await data…
  – OS prolonged and confounded by subsequent (approved) CITs…

• Other endpoints?
  – Immune RECIST since clinical response to immune therapies can manifest after conventional progressive disease (PD) – “pseudoprogression”
  – Tumor growth kinetics as surrogate for response to check-point inhibitors?
Conclusions

• Blinding in CIT trials is a controversial topic

• Important to link the discussion to precise definition of the treatment effect that your clinical trial will estimate (addendum of ICH E9)
  – Facilitates interactions with clinicians, regulators and other stakeholders

• Desire for alternative endpoints… not that easy

• **Remark:** FDA / cross-industry initiative on NPH ongoing, white paper to be expected Q4 2017 / Q1 2018
Doing now what patients need next