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

BBS / PSI CIT event 15-June-2017, Basel

Table of Contents

Table of Contents

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?

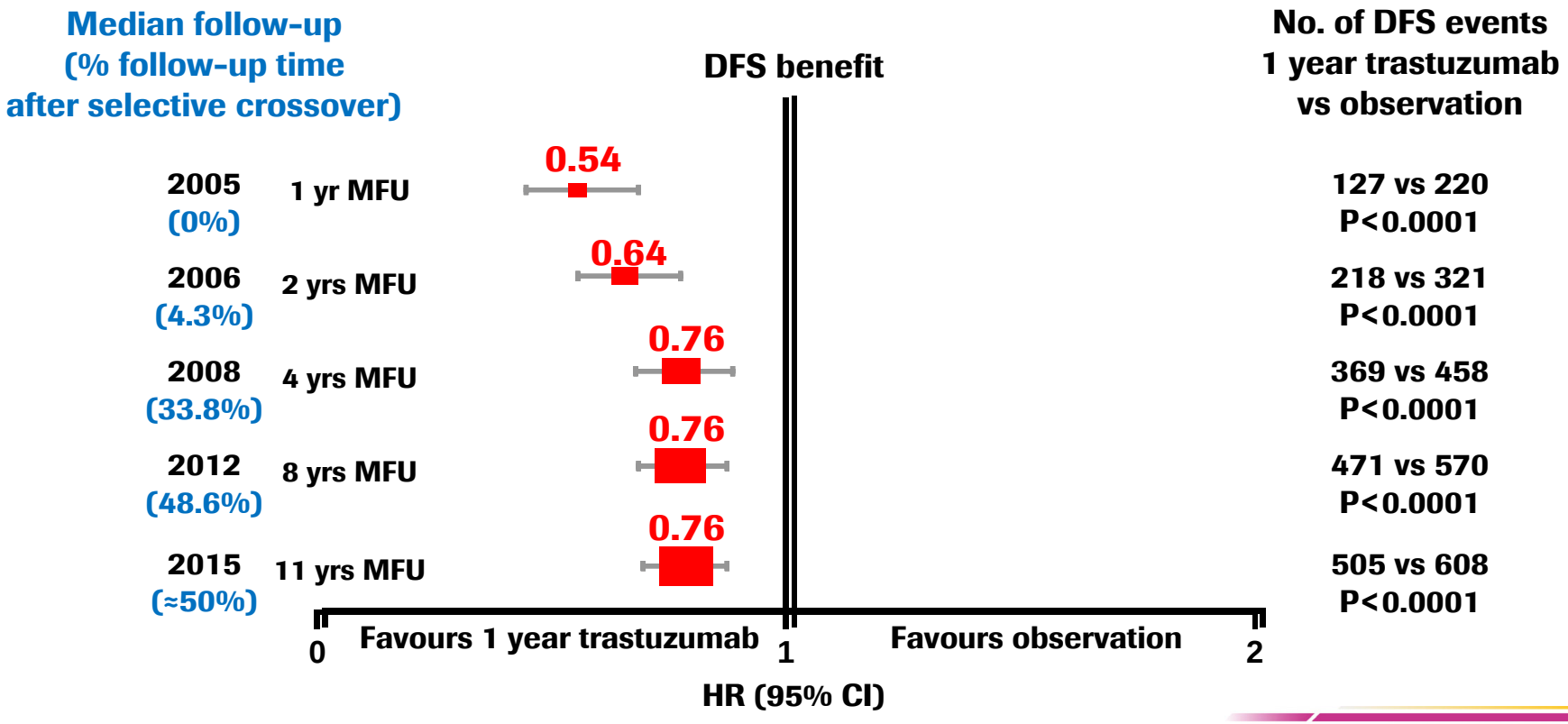
- **Different types of NPH**
  1. Delayed treatment effect
  2. **Diminishing treatment effect (eg treatment switch control arm)**
  3. Crossing Hazard
  4. Long term survivor (“cure” / long-term survival rate)
  5. Subgroup effect: NPH driven by particular subgroup

- 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)



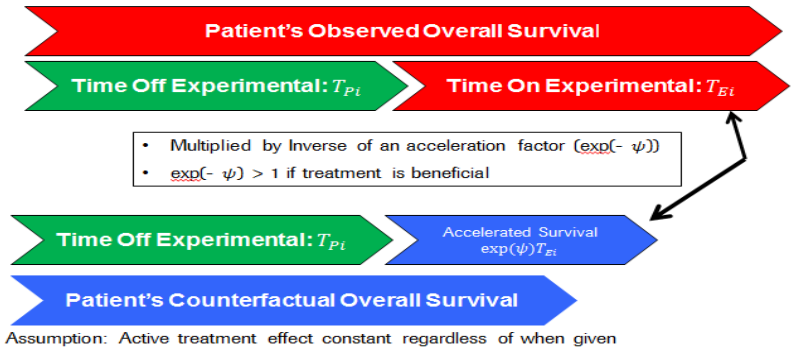
Adjust for switch?

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

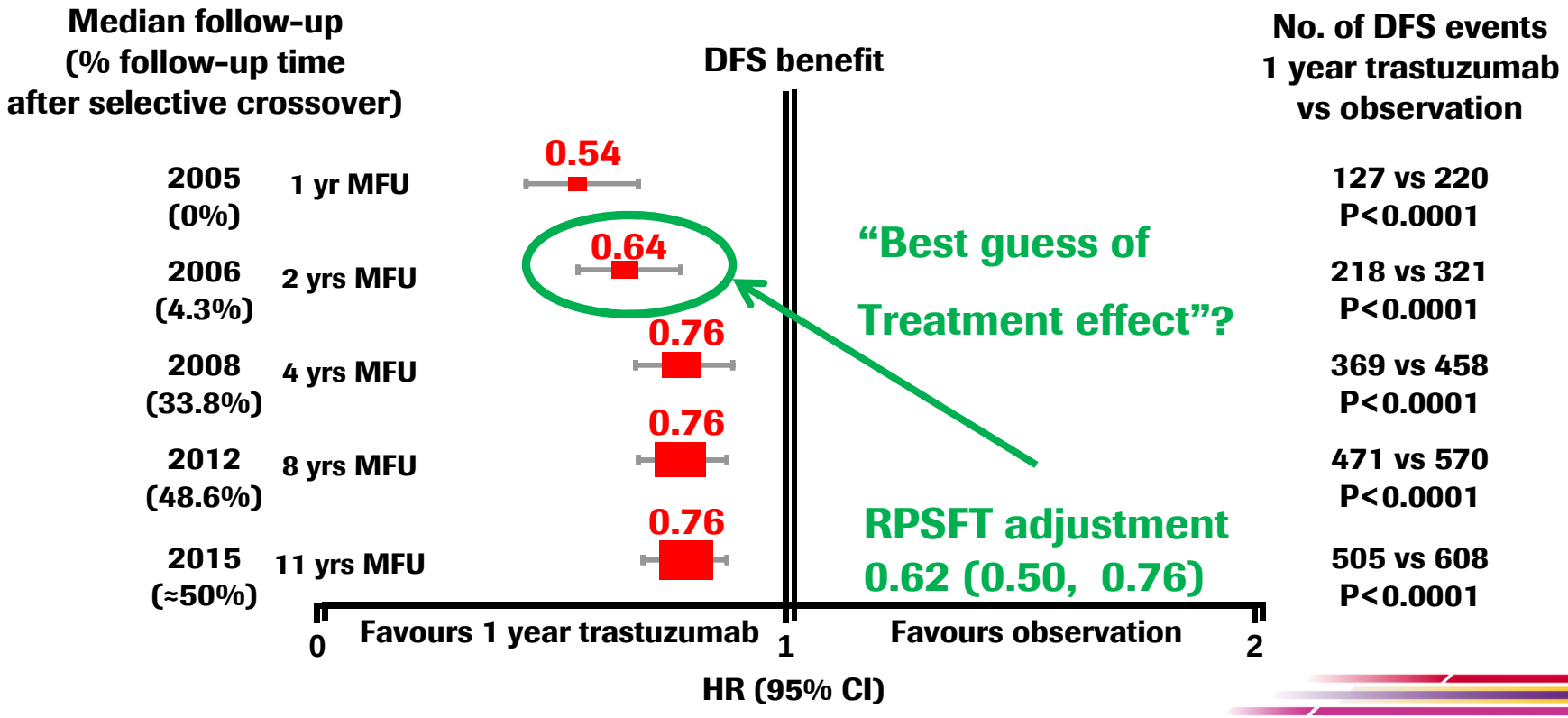
2) **Inverse Probability Censoring Weighting (IPCW)**
   a) Creates a scenario of **missing follow-up data** by censoring the follow-up of each patient at the time of crossover
   b) BUT patients are weighted according to their probability to cross-over
   c) A patient will be assigned a weight of > 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)



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

Effect magnitude of endpoints with CITs

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

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

Scenario	Percentage of switchers (after PFS)	Resulting diminished power
No switch	**0%**	–
Low Switch	**30%**	**-16%**
Medium Switch	**50%**	**-33%**
High Switch	**80%**	**-58%**

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 of Contents

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