Estimands for time to event endpoints in oncology and beyond

Kaspar Rufibach, Hans Ulrich Burger
Methods, Collaboration & Outreach Group (MCO)
Department of Biostatistics, Roche Basel

Context and problem statement

Issues and examples

Role of sensitivity analyses

Discussion

Estimand – a more detailed objective

Choice of estimand involves

1. Population of interest
2. Endpoint of interest
3. Measure of intervention effect

Akacha et al (2015)

Estimand – longitudinal data only?

- Nominal / continuous endpoint measured at one time point without missings: (typically) no ambiguity in definition of estimand
- Longitudinal data for continuous endpoints:
  - **Missing data** likely to occur
  - Missingness typically related to treatment
  - Various analysis methods imply **very different assumptions about missingness**

→ Estimand framework will help to structure all this in protocols

Time-to-event data?

We do not know event date – missing data as well!
(at least when not censored at clinical cut point)

Overall survival (OS):

- Time from registration / randomization to death
- FDA says «event = death due to any reason». Objective
- Counts also deaths unlikely to be due to cancer as events
- How to handle treatment switching if progression-free survival (PFS) is primary endpoint?

→ already for «hard endpoint» OS one can argue about definition of «event» and «censoring»

Analysis of time-to-event data

- Risk set: only contains patients that have been completely observed up to that timepoint

- Survival analysis = analysis of **«completers»** up to a given timepoint
- Standard handling of missing data for such type of data can be considered a **«completer analysis»**
- Realistic assumption in case of only **administrative censoring**
- If censoring not only due to administrative reasons → censoring typically informative → «usual» survival analysis methods biased

- This concerns all time to event analyses, not only in oncology
- Important: Minimize bias by minimizing number of informative censoring or having the same pattern in both groups

Roche

Context and problem statement

Issues and examples

Role of sensitivity analyses

Discussion

Issue 1: Lack of common definition for time-to-event endpoints

- Bellera et al (2013):

  > *Most of these time-to-event endpoints currently **lack standardised definition** enabling a cross comparison of results from different clinical trials.*

- Van Cutsem et al (2005): Randomised trial PETACC 03, colon cancer, primary endpoint disease-free survival (DFS): Results not reported in paper, van Cutsem (2009).
  - Count all secondary primary tumours as event (DFS) → **result significant**
  - Do not count secondary tumors different from colon as event (RF → **result not significant**
  - Again estimand concept could help as it is otherwise not necessarily clear what could be more relevant

- Not a problem for hypothesis test if primary analysis pre-specified. But what about **robustness** of results?

DFS in breast cancer, Hudis et al (2007)

- Primary endpoint for many large adjuvant breast cancer trials
- Typical definition: Randomization to earliest of
  - local
  - regional
  - distant recurrence
  - death
- Often inconsistently defined events:
  - Treatment of contralateral breast cancer
  - Second primary cancers: contralateral? nonbreast? unknown cancers?
  - Death not due to breast cancer

Table 1. Example of Inconsistent Definitions of Disease-Free Survival

Trial	Local/Regional Recurrence	Distant Metastasis	Death From Any Cause	Invasive Contralateral Breast Cancer	Second Primary Invasive Cancer (nonbreast)	Ipsilateral DCIS	Contralateral DCIS	Ipsilateral LCIS	Contralateral LCIS
BIG 1-98[4]	X	X	X	X	X				
MA-17[1]	X	X		X		X	X	X	X
ATAC[2]	X	X	X	X		X	X		
IES[3]	X	X	X	X					
ARNO[5]	X	X		X					

NOTE: Event-free survival used by ARNO.
Abbreviations: DCIS, ductal carcinoma in situ; LCIS, lobular carcinoma in situ; BIG, Breast International Group; MA, National Cancer Institute of Canada Clinical Trials Group MA-17; ATAC, Arimidex, Tamoxifen Alone, or in Combination; IES, Intergroup Exemestane 031; ARNO, Arimidex, Nolvadex 95 Study.

Issue 2: Treatment switching in oncology

- Between 2000 and 2009 debate how to handle patients starting new therapy prior to event of interest
- Two possible approaches:
  1. Censor patients at start of new therapy
  2. **Follow patients up until event and ignore start of new therapy**
- Fleming et al (2009): Approach 2. should be preferred
- Discussion was based on arguments around efficiency (one better than the other) and introduced bias (by censoring them or ignoring further therapy)
- Concept of estimand would have helped the discussion at the time:
  - Do we want to test “time to event” under the assumption “as long as patients stay on study therapy” or under the assumption “irrespective of treatment changes”?
  - Intention-to-treat concept was used to make the point for not censoring patients but was not really powerful. Bias introduced by censoring finally led to the decision towards not censoring

Example: PFS in DLBCL

Diffuse large B-cell lymphoma (DLBCL):

- Accepted endpoint is PFS, registration to earlier of death or progression
- New-anti lymphoma treatment (NALT):
  - Given as 2nd line therapy after progression
  - Without complete response (CR) DLBCL basically a death sentence → sometimes (often?) NALT given before progression to «bring patients to CR»

→ PFS confounded if «too many» NALTs prior to PD?

What do we want to estimate:

1. PFS irrespective of NALT → ignore NALT?
2. «True» time-to-progression without confounding by NALT → censor at NALT prior to PD? Informative censoring?

Context and problem statement

Issues and examples

Role of sensitivity analyses

Discussion

What are meaningful sensitivity analyses?

- Often, **sensitivity analyses** are applied to check robustness and dependency of outcome on analysis assumptions made
- Typically, sensitivity analyses are independent of precise formulation of estimand and estimand changes between primary and sensitivity analysis
  - How should differences be interpreted when a different biological quantity is estimated / tested?
  - Should focus sensitivity analyses as those still estimating / testing the same estimand and keeping other analyses rather as secondary endpoints?
- Estimand should be clearly defined not only for primary but also for sensitivity analyses
- Generally, we should be carefully thinking about the purpose of a sensitivity analysis, what it really adds

Example: Do we / HA know what we want?

Two-arm randomized trial in 2nd line indolent Non-Hodgkin's lymphoma:

- Primary endpoint: PFS
- Submitted sensitivity analysis: Censor patients at last assessment prior to
  1. NALT (purpose: NALT might be indicative of PD)
  2. First missing visit if they had ≥2 missing visits prior to PD or death (purpose: PD might have happened during the time when patient missed visits)

→ Robust assessment of treatment effect. Hazard ratios ≈0.5 consistent

- Estimand not obvious for (1) and (2)
- «Obvious» sensitivity: do not censor but count event

Agency's response: Please provide sensitivity analyses combining (1) & (2)

→ Why? Purpose (guess simply further «robustness» assessment)? Estimand?

→ Substantial programming effort with short turnaround → would be nice to get clear justification of purpose of such analyses, even more when reducing risk of PFS event by half

Context and problem statement

Issues and examples

Role of sensitivity analyses

Discussion

Discussion

- Time-to-event endpoint: **censoring (**before clinical cut) **= missing data**
- Resulting **bias** minimal when number of missing data minimal or similar in both arms
- Depending on how we handle that missing data & also event definition → **estimand not obvious** and should be defined
- For some endpoints in some indications → **heterogeneity in endpoint definitions**. Estimand again not obvious
- **We and Health Authorities** require sensitivity likely to assess «robustness» of primary analysis. Important to put that in estimand framework to clearly understand purpose of such sensitivity analyses

References

- Akacha M, Bretz F, Ohlssen D, Rosenkranz G, Schmidli H. (2015). *Estimands and their role in clinical trials.* Summer issue of the American Statistical Association's Biopharmaceutical Section Report, 22, pp. 1-4.
- Bellera, C.A. et al (2013). *Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project: formal consensus method for the development of guidelines for standardised time-to-event endpoints definitions in cancer clinical trials.* Eur. J. Cancer, 49, 769-781.
- Fleming T.R., et al (2009). Issues in Using Progression-Free Survival When Evaluating Oncology Products. *Journal of Clinical Oncology*. 27(17), 2874-2880.
- Hemmings, R. (2015). *The "estimand" - problem statement.* Presented at PSI "Estimands Discussion Meeting" in February 2015.
- Hudis, C. A., et al (2007). *Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system*. J. Clin. Oncol., 25, 2127-2132.
- Van Cutsem, E. et al (2005). *Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3.* 41st Annual Meeting of the American Society of Clinical Oncology; May 13-17, 2005; Orlando, Fla. Abstract LBA8. http://www.oncologypractice.com/tor/gastrointestinal/50936.html
- Van Cutsem, E. et al (2009). *Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3.* J. Clin. Oncol., 27, 3117-3125.