Estimand framework in Oncology drug development – impact and opportunities

Kaspar Rufibach (Roche), Evgeny Degtyarev (Novartis), Jonathan Siegel (Bayer), Viktoriya Stalbovskaya (Merus), Steven Sun (J&J)
on behalf of the oncology estimand working group

PSI conference London, 4th June 2019

Abstract (will be removed for presentation)

Estimand framework in Oncology drug development – impact and opportunities

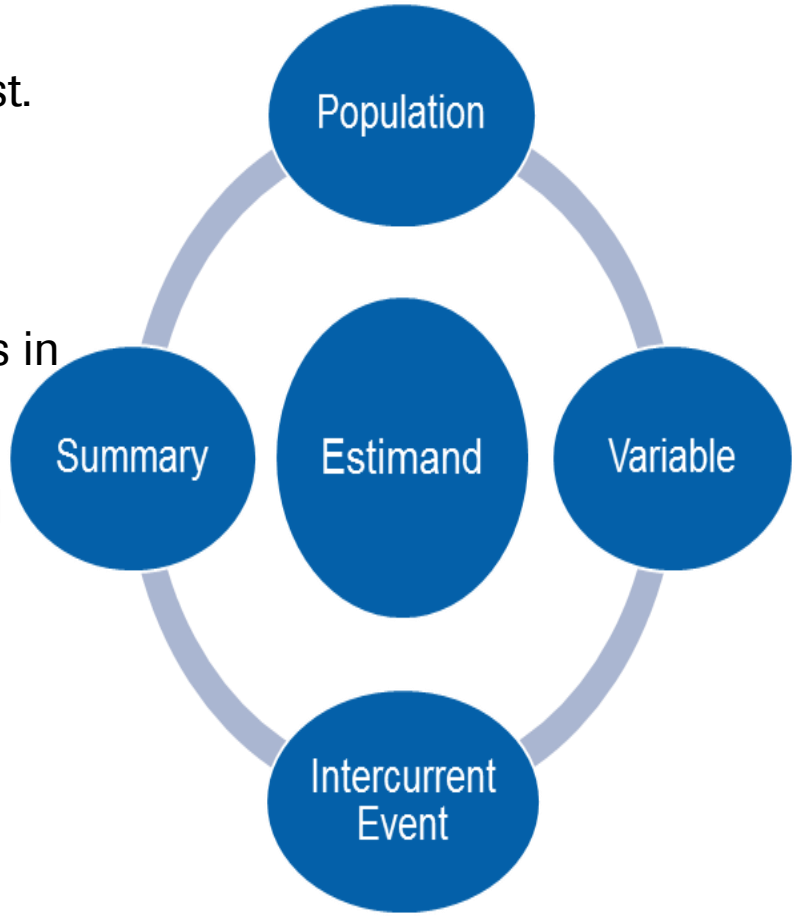
A draft addendum of the ICH E9 guideline on Statistical Principles for Clinical Trials was released in August 2017 and introduced an estimand framework. The new framework aims at aligning trial objectives with design and statistical analyses by requiring a precise definition of the inferential quantity of interest, the estimand. The addendum is anticipated to have a major impact on drug development, as the choice of estimands will drive the trial design, sample size, data collection, trial conduct, analysis, and interpretation.

An industry working group for estimands in oncology was formed in February 2018, with members from 19 companies. The focus areas of the working group are treatment switching, censoring, hematology and solid tumor case studies, and causal estimands in the time-to-event setting. In this talk we will review common estimands of interest and intercurrent events proposed in oncology regulatory guidelines and applications. Several strategies to handle intercurrent events were described in the ICH E9 addendum. These strategies generally interpret intercurrent events as informative or counterfactual outcomes rather than noninformative “missing” data. We will embed those in the time-to-event framework discussing the differences and highlighting the consequences for study design, data collection, analysis and interpretation. Since estimation methods targeting estimands using the proposed strategies often require strong assumptions, the guideline emphasizes sensitivity analyses to justify these. We will discuss sensitivity analyses for key estimands. The concepts will be illustrated using case studies and we will provide recommendations of the industry working group for practical implementation of the estimand framework.

ICH E9 ESTIMAND ADDENDUM AND ONCOLOGY

ICH E9 Addendum and Estimand framework

- ICH E9 (R1) released Aug 2017.
- Precise definition of scientific question of interest.
- Alignment between trial objectives and analysis.
- **Dialogue** between sponsors, regulators, payers, physicians, and patients regarding key questions in clinical trials.
- Framework reflected in several recently released EMA guidelines.
- Concept not restricted to RCTs.



ICH: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

ICH E9 addendum and oncology

- Why this addendum?
  - **Lack of alignment** of trial objectives and effect estimates.
  - Addendum and many (early) publications focus on **longitudinally** measured endpoints, especially with missing data.
- **Time-to-event** (T2E) endpoints?
- Anticipated impact on **oncology** clinical trials?
  - «Oncology» not mentioned in ICH E9 (R1).
  - T2E only marginally.
- Endpoints may reflect time from randomization to
  - death (overall survival, OS),
  - progression or death (progression-free Survival, PFS),
  - progression, start of new therapy, or death (event-free Survival, EFS).

Should we specifically look at T2E endpoints?

- Trials with T2E often long → «allow» more time for intercurrent events to occur.
- Most estimators rely on some sort of censoring:
  - **Random censoring** crucial assumption. Always met?
  - Impact of non-random censoring potentially substantial.
  - Real reason for censoring often not clear.
- Are eCRFs optimally designed to capture:
  - Physician's decision?
  - Patient's decision (i.e. withdrawal consent)?
    separately for withdrawal from treatment and study?

Key questions

- Key **intercurrent** events, endpoints, and estimands in oncology?
- How do five proposed strategies to handle intercurrent events apply to T2E endpoints?
- How can established methods in oncology, e.g.
  - **censoring** schemes or
  - treatment **switching**

  be embedded in addendum framework?
- Sensible **sensitivity** analyses?
- «**Missing data**» often highly informative. What implicit assumptions are we making when simply censoring?
- Quantification of follow-up: often reduction to one single number. What quantity are we interested in with «follow-up»?

Sensitivity and supplementary

- **Sensitivity** analyses:
  - Target same estimand (primary or secondary) under different assumptions.
  - Explore robustness of estimation and data limitations.
- **Supplementary** analyses:
  - Target different estimand than primary.
  - Provide additional insights into the understanding of the treatment effect.

Sensitivity or supplementary – why bother?

- Why bother?
  - No clear estimand targeted.
  - **High number of additional analyses performed,** implying interpretational challenges.

Sensitivity or supplementary – why bother?

3.9.3.4 Sensitivity Analyses

The FL population was the primary population for all efficacy sensitivity analyses.

The following sensitivity analyses for both IRC and investigator-assessed PFS were performed:

- Unstratified log-rank test.
- Re-randomization test of the primary endpoint to assess the sensitivity of the stratified log-rank test to the dynamic randomization procedure. See Kaiser 2012 for details.
- The impact of loss to follow-up was assessed by a worst-case analysis that assigns event outcomes to patients who withdrew prior to disease progression in the obinutuzumab arm at the next scheduled disease assessment date and censored outcomes to patients in the rituximab arm at the last disease assessment date.
- A missed assessment potential impact analysis was performed to assess the robustness of the result of the analysis of PFS. In this analysis, if patients missed an assessment prior to the date of the clinical data cutoff or prior to PD, they were counted as having progressed of the day after their last complete response assessment.
- PFS analyses were repeated with censoring at the initiation of NALT prior to disease progression, to assess potential confounding of the treatment effect estimates by subsequent therapy.
- Patients who discontinued the study treatment for other reasons than disease progression or death were counted as having progressed at the time of discontinuation (event was date of last dose for early treatment discontinuations).
- Patients who died more than 6 months after their last response assessment and showed no sign of progression were censored at the last available response assessment.

Sensitivity or supplementary – why bother?

- Why bother?
  - No clear estimand targeted.
  - **High number of additional analyses performed,** implying interpretational challenges.
  - ICH E9 (R1): «…supplementary analyses should generally be given lower priority than a sensitivity analysis.»
  - Focused on analysis in the past → question should drive analysis!

→ **Less additional analyses** expected **post-addendum** with more **clarity** about purpose and **interpretation.**

ONCOLOGY ESTIMANDS WORKING GROUP

Estimands in Oncology WG

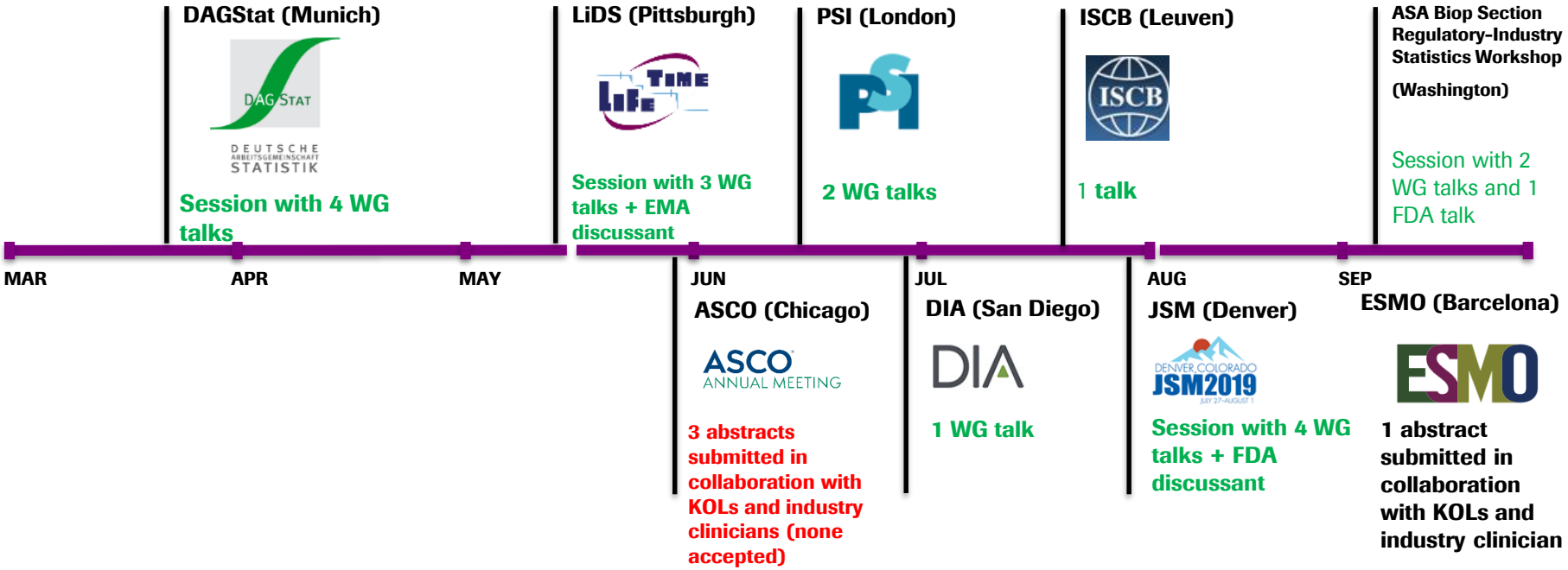
- Initiated and led by Evgeny Degtyarev (Novartis) and Kaspar Rufibach (Roche), first TC Feb 2018.
- Main purpose: ensure **common understanding and consistent definitions** for key estimands in Oncology across industry
- **32 members** (14 from Europe and 18 from US) representing **20 companies**.
- **EFSPI SIG** for Estimands in Oncology (Nov 2018).
- **ASA Biopharm Section Scientific interest WG** (Apr 2019).
- Regular TCs with regulators from EMA, FDA, Japan, China, Taiwan, Canada.

EFSPI SIG: Special Interest Group of European Federation of Statisticians in Pharmaceutical Industry

Estimands in Oncology WG

Communication plan for 2019

- Whitepaper(s) and presentations at statistical and clinical conferences.
- Plans to further engage with Clinical community.



Oncology estimand working group – subteams

1. Causal estimands in T2E setting (Kaspar Rufibach)
   - Bjoern Bornkamp's talk in this session
2. Treatment switching (Viktoriya Stalbovskaya)
3. Censoring mechanisms and their impact on interpretation of estimands (Jonathan Siegel)
4. Case studies in solid tumors (Evgeny Degtyarev)
5. Case studies in hematology (Steven Sun)

EXAMPLES

Sensitivity or supplementary?

- **Investigator-** vs. **independent** radiological assessments for PFS in open-label trials:

Potential bias	Inv-PFS	IRC-PFS
Knowledge of treatment assignment	Inv knows treatment assignment	Performed blinded to treatment assignment
Informative censoring through Inv-PD	N/A	PD called by Inv prior to IRC-PD → scan collection stopped → precludes observation of IRC-PD.

- Questions: Are Inv-PFS and IRC-PFS…
  - … two distinct estimands, with Inv-PD an IC for IRC-PFS?
  - … or two estimators of same estimand, making different assumptions about blinding and censoring?
- WG recommends approach 2 → sensitivity.

Sensitivity or supplementary?

Stratified vs. **unstratified** estimate?

- View 1:
  - Stratified Cox model: Distinct baseline hazard functions for each stratum, common hazard ratio across strata.
  - Unstratified: Identical baseline hazard for each stratum.
  - Same baseline hazard = modeling assumption
    → unstratified sensitivity of primary stratified estimator.
- View 2:
  - Stratified estimator is a conditional estimate → targets conditional estimand → unstratified supplementary of primary stratified estimator.
- WG leaning towards View 2.

Sensitivity or supplementary?

- Conditional effect:
  - Average effect of treatment on individual, i.e. of **moving a subject from untreated to treated**.
  - Estimated from regression coefficient for treatment assignment indicator variable in multiple regression model.
- Marginal effect:
  - Average effect of **moving entire population from untreated to treated**.
  - Unadjusted estimate in RCT.

Estimand	Linear regression	Logistic regression	Cox regression	Aalen additive model
Unadjusted	Marginal	Marginal	Marginal	Marginal
Covariate-adjusted	Effect collapsible, i.e. marginal = conditional	Conditional	Conditional	Effect collapsible, i.e. marginal = conditional

- Do not routinely run adjusted and unadjusted analysis → they may target **different estimand**! One supplementary of the other.
- First define estimand, then estimator.
- FDA guideline on adjustment for covariates for continuous outcomes.

Non-proportional hazards

- Unlikely that a **single estimand** adequately describes treatment effect under NPH.
- Specify single estimand: reduced power unless you made correct assumption.
- Specify multiple estimands:
  - For example through **weighted log-rank combination tests** → optimize power, but no clear estimand?
  - Trade-off variability against bias.
- What estimand corresponds to hazard ratio in Cox regression?
- These points **surfaced and gained in importance** through estimand debate.

Opportunities for statisticians

- Embrace and develop **new methods** for T2E endpoints:
  - Competing risk and multi-state models.
  - Causal methods to deal with confounding variables or post-randomization events.
  - Replace naive (often misleading!) analyses through e.g. analyses based on principal stratification.
  - How to deal with non-proportional hazards?
- Evaluate **bias – variance tradeoffs**, offer alternative solutions:
  - censoring,
  - Inv vs. IRC PFS,
  - non-proportional hazards,
  - …

Opportunities for statisticians

- If well justified, more **flexible approaches** might become acceptable to regulators:
  - Hypothetical OS estimand with X-over after PFS: https://www.ema.europa.eu/en/adjustment-cross-over-estimating-effects-oncology-trials
  - EMA already updated diabetes and AD guideline.
- Clarify / align questions industry and/or regulators have handled heterogeneously in the past:
  - Inv-PFS vs. IRC-PFS.
  - Stratified vs. unstratified.
  - Start of new therapy in absence of PD.
  - …
- Make dialogue which data needs to be collected easier.
- Involve broader teams **early** in trial planning in estimand discussion.

Conclusions

- Addendum has potential to change way we design and analyze trials, and collect data. Or already did!
- Addendum will bring more **transparency** around
  - connection of trial objective to estimand and estimator,
  - bias-variance trade-off of a given estimator,
  - handling of «missing» data,
  - interpretation of trial results and added value of drugs.
- Opportunity for **structured dialogue** between all **stakeholders**.
- Reduce overall number of (unfocused) analyses.
- **Leadership opportunity** for statisticians that are able to connect clinical to statistical questions.

Thank you for your attention.

BACKUP

Estimand framework and possible strategies to handle intercurrent events

- **Treatment Policy**: occurrence of intercurrent event is irrelevant
- **Composite**: intercurrent event is considered component of the variable
- **Hypothetical**: a hypothetical scenario is envisaged in which the intercurrent event would not occur
- **Principal stratification**: population is defined by a patient's potential intercurrent events on either or both treatments
- **While on treatment**: response to treatment prior to the occurrence of the intercurrent event is of interest

To censor or not to censor?

- Handling of **alternative anticancer treatment (intercurrent event) in PFS analysis prior to observed progression or death** ?

- Cheson et al (2007, 2572 citations on 5th March) and FDA guideline:
  - «... such patients **should be censored**...»
  - «imputes» PFS for these patients using those who benefitted from treatment and had longer follow-up
  - potential risk of non-random censoring
  - **Hypothetical** estimand
    - in T2E setting censoring can also be applied for while on treatment estimand

To censor or not to censor?

- Fleming et al (2009, 105 citations), and also EMA guideline:
  - «...patients **should not be censored** at the time other treatments are initiated when analyzing the PFS end point…»
  - **Treatment policy** estimand
- «This induces strongly dependent censoring because the true time to progression for that patient is replaced by the true time to progression of other patients who also were free of progression at month *x*, but did not need other treatments at that time.»
- Treatment policy estimand favoured to avoid non-random censoring
- Change the question of interest because of challenge in analysis?
- Thinking should be reversed:
  Trial objective → estimand → estimator.

References

Cheson, B. D., Pfistner, B., Juweid, M. E., Gascoyne, R. D., Specht, L., Horning, S. J., Coiffier, B., Fisher, R. I., Hagenbeek, A., Zucca, E., Rosen, S. T., Stroobants, S., Lister, T. A., Hoppe, R. T., Dreyling, M., Tobinai, K., Vose, J. M., Connors, J. M., Federico, M. and Diehl, V. (2007). *Revised response criteria for malignant lymphoma*. J. Clin. Oncol. 25 579--586.

Fleming, T. R., Rothmann, M. D. and Lu, H. L. (2009). *Issues in using progression-free survival when evaluating oncology products*. J. Clin. Oncol. 27 2874--2880.

ICH E9 working group (2014). *E9(R1): Addendum to Statistical Principles for Clinical Trials on Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials*. link

ICH E9 working group (2017). *ICH E9 (R1): addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials*. Link

Rufibach, K. (2019). *Treatment effect quantification for time-to-event endpoints -- estimands, analysis strategies, and beyond*. Pharmaceutical Statistics, 18, 144-164. doi

Doing now what patients need next