Survival analysis for AdVerse events with VarYing follow-up times – The SAVVY Project

Regina Stegherr[1]

[1] This work was conducted at the Institute of Statistics, Ulm University, Germany, and is not in any way related to my current work at the "Institut für Klinische Pharmakologie und Toxikologie, Pharmakovigilanz, und Beratungszentrum für Embryonaltoxikologie, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin", Germany

25.08.2022
Paul Martini Preis 2022

regina.stegherr@charite.de

Thanks to...

The SAVVY project group

- Academic leads: Jan Beyersmann (Ulm), Tim Friede (Göttingen) and Claudia Schmoor (Freiburg)
- Steering Committee: Valentine Jehl (Novartis), Friedhelm Leverkus (Pfizer), Kaspar Rufibach (Roche) and the academic leads
- Participating companies: Bayer, Boehringer Ingelheim, BMS, Janssen, Lilly, Merck, Novartis, Pfizer, Roche

Survival analysis for AdVerse events with VarYing follow-up times - SAVVY

- **Safety in terms of adverse events (AEs)** is a relevant aspect of risk-benefit assessent of therapies (Unkel et al., 2019).
- In analyses of AEs (of a certain kind), observation may be precluded by **death, progression** or some other competing event. Moreover, recording of AEs is limited to a restricted period of time (**censoring**) and varying follow-up times (Allignol et al., 2016).
- Overall aim: Improve reporting of AEs through the use of survival techniques appropriately dealing with varying follow-up times and competing events.
- **Empirical study**: Investigate in several randomized controlled trials whether the use of different estimators for analyses of AEs leads to different conclusions about therapies' safety.
- Comparison of **commonly** used (but biased) **estimators quantifying the AE probability** to estimators **accounting for competing events** in time-to-event studies and also compare safety comparisons between treatment groups.

Survival analysis for AdVerse events with VarYing follow-up times - SAVVY

Four articles evolved from the project

- **Methodological considerations and simulations**: [1] Stegherr, R., Schmoor, C., Lübbert, M., Friede, T., & Beyersmann, J. (2021). Estimating and comparing adverse event probabilities in the presence of varying follow-up times and competing events. *Pharmaceutical Statistics*, 20(6), 1125-1146.
- **Statistical concept of the empirical study**: [2] Stegherr, R., Beyersmann, J., Jehl, V., Rufibach, K., et al. (2021). Survival analysis for AdVerse events with VarYing follow-up times (SAVVY): Rationale and statistical concept of a meta-analytic study. *Biometrical Journal*, 63(3), 650-670.
- **Results for probability estimators**: [3] Stegherr, R., Schmoor, C., Beyersmann, J., Rufibach, K., et al. (2021). Survival analysis for AdVerse events with VarYing follow-up times (SAVVY)–estimation of adverse event risks. *Trials*, 22(1), 1-13.
- **Results for group comparisons**: [4] Rufibach, K.*, Stegherr, R.*, Schmoor, C., Jehl, et al. (2020). Survival analysis for AdVerse events with VarYing follow-up times (SAVVY)–comparison of adverse event risks in randomized controlled trials. arXiv preprint arXiv:2008.07881. (* both authors contributed equally to this work.)

Estimating AE probabilities: Commonly used but biased methods

- **Incidence proportion**: $\frac{\text{\# AEs in } [0, \tau]}{\text{\# patients}}$
  - Usually only calculated at the end of follow-up $\Rightarrow$ Assumes identical follow-up times in all patients
  - **Underestimation** of AE probability in presence of censoring
- **Incidence density**: $\mathrm{ID}(\tau) = \frac{\text{\# AE in } [0, \tau]}{\text{patient-time at risk restricted by } \tau}$
  - Assumption of constant hazards
  - Estimator of hazard rate $\Rightarrow$ probability scale requires transformation: $1 - \exp\left(-\mathrm{ID}(\tau) \cdot \tau\right)$
  - Parametric version of 1-Kaplan-Meier
- **1-Kaplan-Meier**: competing events censored at their event time
  - **Overestimation** of AE probability in presence of competing events
  - About 50% of all Kaplan-Meier curves ignore competing risks (van Walraven et al. 2016, Schumacher et al. 2016)

Estimating AE probabilities: Commonly used but biased methods

- **Incidence proportion**: $\frac{\text{\# AEs in } [0, \tau]}{\text{\# patients}}$
  - Usually only calculated at the end of follow-up $\Rightarrow$ Assumes identical follow-up times in all patients
  - **Underestimation** of AE probability in presence of censoring
- **Incidence density**: $\text{ID}(\tau) = \frac{\text{\# AE in } [0, \tau]}{\text{patient-time at risk restricted by } \tau}$
  - Assumption of constant hazards
  - Estimator of hazard rate $\Rightarrow$ probability scale requires transformation: $1 - \exp\left(-\text{ID}(\tau) \cdot \tau\right)$
  - Parametric version of 1-Kaplan-Meier
- **1-Kaplan-Meier**: competing events censored at their event time
  - **Overestimation** of AE probability in presence of competing events
  - About 50% of all Kaplan-Meier curves ignore competing risks (van Walraven et al. 2016, Schumacher et al. 2016)

Estimating AE probabilities: Commonly used but biased methods

- **Incidence proportion**: $\frac{\text{\# AEs in } [0, \tau]}{\text{\# patients}}$
  - Usually only calculated at the end of follow-up $\Rightarrow$ Assumes identical follow-up times in all patients
  - **Underestimation** of AE probability in presence of censoring
- **Incidence density**: $\text{ID}(\tau) = \frac{\text{\# AE in } [0, \tau]}{\text{patient-time at risk restricted by } \tau}$
  - Assumption of constant hazards
  - Estimator of hazard rate $\Rightarrow$ probability scale requires transformation: $1 - \exp\left(-\text{ID}(\tau) \cdot \tau\right)$
  - Parametric version of 1-Kaplan-Meier
- **1-Kaplan-Meier**: competing events censored at their event time
  - **Overestimation** of AE probability in presence of competing events
  - About 50% of all Kaplan-Meier curves ignore competing risks (van Walraven et al. 2016, Schumacher et al. 2016)

Estimating AE probabilities: Alternative, underused approaches

- **Aalen-Johansen estimator**: $\mathrm{CIF}(\tau) = \sum_{u \in (0,\tau]} \prod_{v \in (0,u)} \left(1 - \Delta\hat{\Lambda}(v) - \Delta\hat{\overline{\Lambda}}(v)\right) \Delta\hat{\Lambda}(u)$
  - **Gold-standard**: accounts for censoring and competing events and is not restricted to constant hazards (non-parametric)
  - Generalizes the Kaplan-Meier estimator to multiple event types
- **Probability transform of the incidence density accounting for competing events** (parametric version of Aalen-Johansen): $\frac{\mathrm{ID}(\tau)}{\mathrm{ID}(\tau) + \overline{\mathrm{ID}}(\tau)} \left(1 - \exp(-\tau \cdot [\mathrm{ID}(\tau) + \overline{\mathrm{ID}}(\tau)])\right)$

  with $\overline{\mathrm{ID}}(\tau) = \frac{\#\text{ competing event in } [0,\tau]}{\text{patient-time at risk restricted by } \tau}$

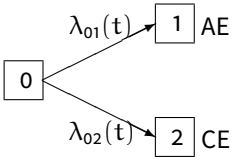
  - Assumption of **constant hazards** for both (AE and competing event) hazards
  - In literature about incidence densities often neglected (e.g. book 'analysis of incidence rates' by Cummings, 2019)

Estimating AE probabilities: Alternative, underused approaches

- **Aalen-Johansen estimator**: $\mathrm{CIF}(\tau) = \sum_{u \in (0,\tau]} \prod_{v \in (0,u)} \left(1 - \Delta\hat{\Lambda}(v) - \Delta\hat{\overline{\Lambda}}(v)\right) \Delta\hat{\Lambda}(u)$
  - **Gold-standard**: accounts for censoring and competing events and is not restricted to constant hazards (non-parametric)
  - Generalizes the Kaplan-Meier estimator to multiple event types
- **Probability transform of the incidence density accounting for competing events** (parametric version of Aalen-Johansen): $\frac{\mathrm{ID}(\tau)}{\mathrm{ID}(\tau) + \overline{\mathrm{ID}}(\tau)} \left(1 - \exp(-\tau \cdot [\mathrm{ID}(\tau) + \overline{\mathrm{ID}}(\tau)])\right)$

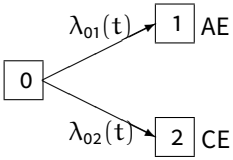
  with $\overline{\mathrm{ID}}(\tau) = \frac{\#\text{ competing event in } [0, \tau]}{\text{patient-time at risk restricted by } \tau}$
  - Assumption of **constant hazards** for both (AE and competing event) hazards
  - In literature about incidence densities often neglected (e.g. book 'analysis of incidence rates' by Cummings, 2019)

Definition of the competing event



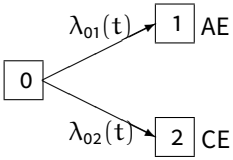
- Time-to-1st-event and type-of-1st-event
- **Adverse event (AE):** Event of interest
- Two possible definitions of a competing event (CE):
  - **Death only**: death without prior AE, i.e., events after which an AE can definitely not occur any more
  - **All events**: death, loss to follow up, withdrawal of consent, treatment discontinuation, and progression, i.e., competing events after which an AE in principle still could occur, but is not observed due to premature end of follow-up
- **Censoring:** designated end of follow-up reached without having an AE or a competing event as defined above; administrative not triggered by course of disease

Definition of the competing event



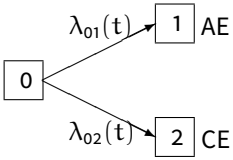
- Time-to-1st-event and type-of-1st-event
- **Adverse event (AE):** Event of interest
- Two possible definitions of a competing event (CE):
  - **Death only**: death without prior AE, i.e., events after which an AE can definitely not occur any more
  - **All events**: death, loss to follow up, withdrawal of consent, treatment discontinuation, and progression, i.e., competing events after which an AE in principle still could occur, but is not observed due to premature end of follow-up
- **Censoring:** designated end of follow-up reached without having an AE or a competing event as defined above; administrative not triggered by course of disease

Definition of the competing event



- Time-to-1st-event and type-of-1st-event
- **Adverse event (AE):** Event of interest
- Two possible definitions of a competing event (CE):
  - **Death only**: death without prior AE, i.e., events after which an AE can definitely not occur any more
  - **All events**: death, loss to follow up, withdrawal of consent, treatment discontinuation, and progression, i.e., competing events after which an AE in principle still could occur, but is not observed due to premature end of follow-up
- **Censoring:** designated end of follow-up reached without having an AE or a competing event as defined above; administrative not triggered by course of disease

Definition of the competing event



- Time-to-1st-event and type-of-1st-event
- **Adverse event (AE):** Event of interest
- Two possible definitions of a competing event (CE):
  - **Death only**: death without prior AE, i.e., events after which an AE can definitely not occur any more
  - **All events**: death, loss to follow up, withdrawal of consent, treatment discontinuation, and progression, i.e., competing events after which an AE in principle still could occur, but is not observed due to premature end of follow-up
- **Censoring:** designated end of follow-up reached without having an AE or a competing event as defined above; administrative not triggered by course of disease

Group comparisons and follow-up times

- **Risk difference** or **relative risk** of incidence proportions may be misleading
  - Comparing two quantities that both underestimate the AE probability
  - Comparing two quantities **evaluated at different follow-up times**, i.e., largest observed event time in experimental treatment group E τ_E may be greater/smaller than largest observed event time in comparison group C τ_C (Incidence proportion only calculated at the end of follow-up, Bender et al., 2016) (referred to as maximum follow-up time)
- Evaluate estimators at $\tau = \min(\tau_E, \tau_C)$ (referred to common maximum follow-up time)
- As estimators (e.g. Kaplan-Meier) at the end of follow-up may have **larger variability** due to small numbers still at risk (Pocock et al. 2002):
  - Evaluate estimators at **earlier time point when more patients are still at risk**
  - Evaluate estimators at $\tilde{\tau} = \min(\tilde{\tau}_E, \tilde{\tau}_C)$, with $\tilde{\tau}_E(p)$ and $\tilde{\tau}_C(p)$ defined as event time when $p \cdot 100\%$ of all patients in group E and group C, respectively, are still at risk , e.g., $p = 0.9$ (P90), $p = 0.6$ (P60) and $p = 0.3$ (P30)

Group comparisons and follow-up times

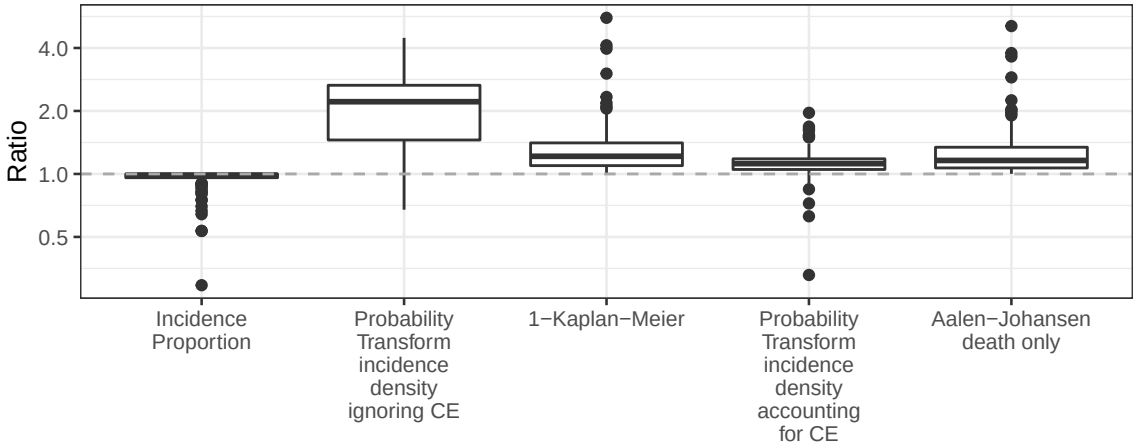
- **Risk difference** or **relative risk** of incidence proportions may be misleading
  - Comparing two quantities that both underestimate the AE probability
  - Comparing two quantities **evaluated at different follow-up times**, i.e., largest observed event time in experimental treatment group E τ_E may be greater/smaller than largest observed event time in comparison group C τ_C (Incidence proportion only calculated at the end of follow-up, Bender et al., 2016) (referred to as maximum follow-up time)
- Evaluate estimators at $\tau = \min(\tau_E, \tau_C)$ (referred to common maximum follow-up time)
- As estimators (e.g. Kaplan-Meier) at the end of follow-up may have **larger variability** due to small numbers still at risk (Pocock et al. 2002):
  - Evaluate estimators at **earlier time point when more patients are still at risk**
  - Evaluate estimators at $\tilde{\tau} = \min(\tilde{\tau}_E, \tilde{\tau}_C)$, with $\tilde{\tau}_E(p)$ and $\tilde{\tau}_C(p)$ defined as event time when $p \cdot 100\%$ of all patients in group E and group C, respectively, are still at risk , e.g., $p = 0.9$ (P90), $p = 0.6$ (P60) and $p = 0.3$ (P30)

Empirical Study

- Statistical Analysis Plan can be found in Stegherr et al. (2021, BiomJ)
- Only **aggregated** data shared with the project collaborators: Trial level analyses ran within the sponsor company / organization using SAS (and R) code provided by the project collaborators $\Rightarrow$ **no release of individual patient data was required**.
- **Pilot study** to **develop SAS macros**, to **assess feasibility** of macros and output data structure, to **check output dataset** whether they contain all necessary information and to **train meta-analysis** and obtain early results (3 partners providing 5 trials and a total of 62 types AEs (range 2 - 51 per trial))
- **Main study**: 10 participating organizations contributing 17 trials including 186 types of AEs

Boxplots of the ratio estimator of interest/gold-standard Aalen-Johansen estimator

AE probability in group E at maximum follow-up time



Meta-analysis

AE probability in group E

Observed data: estimator of log-ratio (log(estimator/gold-standard Aalen-Johansen)) $\hat{\theta}_k$ with bootstrapped variance $\hat{\sigma}^2_k$, $k = 1, ..., 186$ types of AEs

- Normal-normal hierarchical model (NNHM): $\hat{\theta}_j|\theta_j \sim N(\theta_j, \sigma^2_j)$, $\theta_j|\theta, \rho \sim N(\theta, \rho^2)$, $j = 1, ..., K$
- Interpretation of estimate $\hat{\theta}$ (intercept): $\exp(\hat{\theta})$ corresponds to the estimated average ratio

FU time	IP	Prob Trans ID	1-KM	Prob Trans ID CE	AJE (death)
maximum	0.972	2.097	1.214	1.130	1.170

- Univariable and multivariable meta-regression to see what drives the size of the bias; Input variables: value of gold-standard estimator, proportion of censoring, proportion of competing events, maximal follow-up time in experimental group

Meta-analysis

AE probability in group E

Observed data: estimator of log-ratio (log(estimator/gold-standard Aalen-Johansen)) $\hat{\theta}_k$ with bootstrapped variance $\hat{\sigma}^2_k$, $k = 1, ..., 186$ types of AEs

- Normal-normal hierarchical model (NNHM): $\hat{\theta}_j|\theta_j \sim N(\theta_j, \sigma^2_j)$, $\theta_j|\theta, \rho \sim N(\theta, \rho^2)$, $j = 1, ..., K$
- Interpretation of estimate $\hat{\theta}$ (intercept): $\exp(\hat{\theta})$ corresponds to the estimated average ratio

FU time	IP	Prob Trans ID	1-KM	Prob Trans ID CE	AJE (death)
maximum	0.972	2.097	1.214	1.130	1.170

- Univariable and multivariable meta-regression to see what drives the size of the bias; Input variables: value of gold-standard estimator, proportion of censoring, proportion of competing events, maximal follow-up time in experimental group

Summary

- "common" and "very common" AEs overrepresented and most studies from oncology ⇒ **The empirical study illustrates possible biases but the results can not be generalized;**
- Simulations in line with the results of the empirical study (Stegherr et al. 2021, Pharm Stat)
- Also considered in the empirical study: Earlier follow-up times, frequency categories of probability estimates, group comparisons, Cox model vs ratio of incidence densities, and composite endpoint
- Choice of estimator crucial for estimation of AE probability and for group comparisons
- If time-to-event methods are considered to account for censoring instead of simply using the incidence proportion, it is important to correctly analyze competing events (**Kaplan-Meier must not be used**)
- Ignoring competing events worse than assuming simple constant hazards model
- **Best choice**: Always use Aalen-Johansen estimator to estimate AE probability and use RR calculated with Aalen-Johansen and HR from Cox model for **all** types of events that are considered for group comparisons

Summary

- “common” and “very common” AEs overrepresented and most studies from oncology ⇒ **The empirical study illustrates possible biases but the results can not be generalized;**
- Simulations in line with the results of the empirical study (Stegherr et al. 2021, Pharm Stat)
- Also considered in the empirical study: Earlier follow-up times, frequency categories of probability estimates, group comparisons, Cox model vs ratio of incidence densities, and composite endpoint
- Choice of estimator crucial for estimation of AE probability and for group comparisons
- If time-to-event methods are considered to account for censoring instead of simply using the incidence proportion, it is important to correctly analyze competing events (**Kaplan-Meier must not be used**)
- Ignoring competing events worse than assuming simple constant hazards model
- **Best choice**: Always use Aalen-Johansen estimator to estimate AE probability and use RR calculated with Aalen-Johansen and HR from Cox model for **all** types of events that are considered for group comparisons

References

- Stegherr, R., Schmoor, C., Beyersmann, J., Rufibach, K., Jehl, V., Brückner, A., Eisele, L., Künzel, K., Kupas, K., Langer, F., Leverkus, F., Loos, A., Norenberg, C., Voss, F. and Friede, T. (2020). Survival analysis for AdVerse events with VarYing follow-up times (SAVVY) — estimation of adverse event risks. *Trials*, 22(1), 1-13.
- Rufibach, K., Stegherr, R., Schmoor, C., Jehl, V., Allignol, A., Boeckenhoff, A., Dunger-Baldauf, C., Eisele, L., Künzel, T., Kupas, K., Leverkus, F., Trampisch, M., Zhao, Y., Friede, T. and Beyersmann, J. (2020) Survival analysis for AdVerse events with VarYing follow-up times (SAVVY) – comparison of adverse event risks in randomized controlled trials. arXiv:2008.07881 (submitted).
- Stegherr, R., Beyersmann, J., Jehl, V., Rufibach, K., Leverkus, F., Schmoor, C. and Friede, T. (2020). Survival analysis for AdVerse events with VarYing follow-up times (SAVVY): Rationale and statistical concept of a meta-analytic study. *Biometrical Journal*, *Biometrical Journal*, 63(3), 650-670.
- Stegherr, R., Schmoor, C., Lübbert, M., Friede, T. and Beyersmann, J. (2020). Estimating and comparing adverse event probabilities in the presence of varying follow-up times and competing events. *Pharmaceutical Statistics*, 20(6), 1125-1146.
- Allignol, A., Beyersmann, J. and Schmoor, C. (2016). Statistical issues in the analysis of adverse events in time-to-event data. *Pharmaceutical Statistics* **15**, 297–305.
- Bender, R., and Beckmann, L. (2019). Limitations of the incidence density ratio as approximation of the hazard ratio. *Trials*, **20**(1), 485.
- Pocock, S. J., Clayton, T. C. and Altman, D. G. (2002). Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *The Lancet* **359**, 1686–1689.
- Schumacher, M., Ohneberg, K., and Beyersmann, J. (2016). Competing risk bias was common in a prominent medical journal. *Journal of clinical epidemiology*, **80**, 135-136.
- Unkel, S., Amiri, M., Benda, N., Beyersmann, J., Knoerzer, D., Kupas, K., Langer, F., Leverkus, F., Loos, A., Ose, C., Proctor, T., Schmoor, C., Schwenke, C., Skipka, G., Unnebrink, K., Voss, F. and Friede, T. (2019). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharmaceutical Statistics* **18**, 165–183.