Vaccine Efficacy waning estimation and extrapolation using causal inference

Andrea Callegaro[1] , Jyoti Soni[2]

1.GSK; Statistical and Data Science Innovation Hub (SDSIH), Belgium
2. GSK; Biostatistics, Belgium

PSI, Wembley Stadium, United Kingdom (09JUN2025)

Conflict of interest

- GlaxoSmithKline was the funding source for all costs associated with this work.
- AC, JS are employees of the GSK group of companies and own shares in the GSK group of companies.

Content of the presentation

- Objective of the exploration and overview of the hypothetical case study
- Causal inference: Parametric G-computation concept
- Application to the hypothetical case study
- Causal Hazard Ratio sensitivity analysis
- Competing risk analysis
- Conclusions

Objective of the exploration and overview of the hypothetical case study

Objectives:

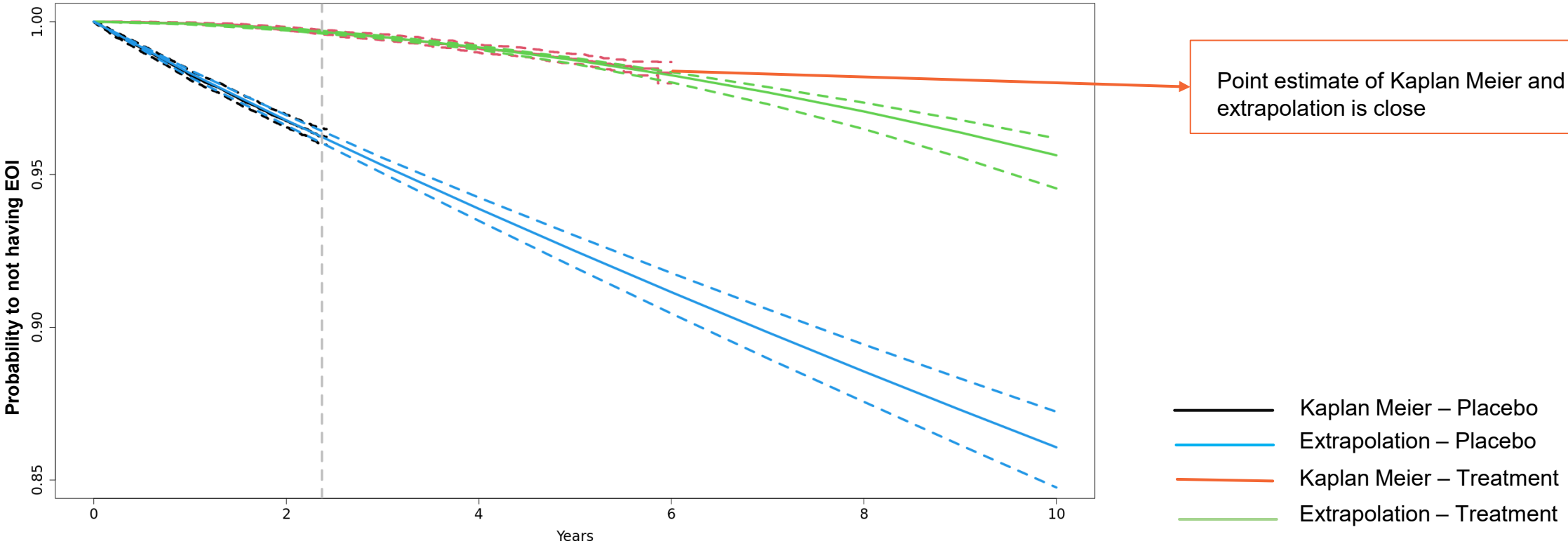
- To extrapolate the long-term Vaccine Efficacy (VE) against Event of Interest (EoI)

Specificity of study design of the hypothetical case study

- Large Phase III, randomized, placebo controlled study with vaccination and long term follow-up
- Enrollment was randomised based on prognostic factors X1 (continuous), X2 and X3 (categorical)
- Provided the efficacy is established, placebo arm participants were vaccinated and stopped from follow-up
- Data available from Placebo is for maximum of 2.5 years and from the vaccinated group for maximum of 6.25 years

Extrapolation using parametric survival models

Why parametric survival models? Because they are needed for extrapolation!



This figure shows

- Kaplan-Meier curve by treatment group and extrapolation using parametric survival model (Flexible Weibull)

Extrapolation is very close to Kaplan Meier

Causal Inference: Potential outcomes and Estimands

Potential outcomes T_i^a represent the time from origin to event for participant i that would have occurred under exposure a.

Estimands

Hazard ratio (HR) - Vaccine effect at time t

$$HR(t) = \frac{lim_{h\to 0}(t \leq T^1 < t+h | T^1 > t)}{lim_{h\to 0}(t \leq T^0 < t+h | T^0 > t)} = \frac{\lambda^1(t)}{\lambda^0(t)}$$

WARNING: *problematic causal interpretation* because it suffers from so-called built-in selection bias, over time (Hernan, 2010).

Cumulative incidence ratio (CIR) – Cumulative vaccine effect over time 0 up to t

$F^a(t) = 1 - S^a(t) = P(T^a < t)$: Risk (cumulative probability) of experiencing the event by time t under exposure plan a,

$$CIR(t) = \frac{F^1(t)}{F^0(t)}$$

Causal interpretation but less informative for booster decision-making (it is a cumulative effect).

Potential outcomes and Estimands – Continues

Causal Hazard ratio (CHR) (Martinussen et al., 2020)

Instantaneous risk at time *t* on treatment versus control for the **principal stratum** of individuals who would have survived up to time t, no matter what treatment.

$$CHR(t) = \frac{lim_{h\to 0}(t \le T^1 < t+h | T^0 > t, T^1 > t)}{lim_{h\to 0}(t \le T^0 < t+h | T^0 > t, T^1 > t)}$$

WARNING. CHR *"is of theoretical interest only, since they rely on assumptions that cannot be empirically evaluated"* (Martinussen et al., 2020).

Axelrod et al, (2022) proposed a sensitivity analysis based on frailty models

Estimators: parametric G-computation

G-computation is an efficient estimation technique for causal inference adjusting for covariates (Book - “What if”)

Under the classical assumptions of exchangeability, positivity, and consistency (and non informative censoring), the following parametric G-computation approach can be used (Hernan, 2010; “What if” book pag. 230) in following steps:-

1. Fit a *flexible parametric survival model* to predict the conditional survival function given exposure, baseline covariates, and time of follow-up.
2. Predict the survival at time *t* for each subject both under exposure and under no exposure, regardless of the subject's exposure status.
3. Separately average the conditional survivals under exposure and under no exposure, over all subjects (*standardized survival*)

$$S(t|A=a) = E[S(t|A=a,X)] = \frac{1}{N}\sum_{i=1}^{N} S(t|A=a,X=x_i)$$

It can be shown (Rutherford et al. (2020), Appendix I) that the marginal hazard of the standardized survival can be calculated as

$$h(t|A=a) = \frac{\sum_{i=1}^{N} S(t|A=a,X=x_i)h(t|A=a,X=x_i)}{\sum_{i=1}^{N} S(t|A=a,X=x_i)}$$

These estimators can be computed using the R package flexsurv.

Competing risks background

- In failure-time settings, a competing event is any event that makes it impossible for the event of interest to occur.

$$T = \min(T_1, T_2) \text{ overall event-time, where } T_j \text{ is the time for event } j = 1, 2$$

$$F_1^{\text{sub}}(t) = P(T_1 \leq t) \quad \text{subdistribution function}$$

$$h_1^{\text{sub}}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_1 < t + \Delta t \mid T_1 \geq t)}{\Delta t} \quad \text{subdistribution hazard}$$

$$h_1^{\text{CS}}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, j = 1 \mid T \geq t)}{\Delta t} \quad \text{cause-specific hazard}$$

* in the cause-specific hazard, individuals are no longer in the risk set after an *event of any type*
in the subdistribution hazard, individuals leave the risk set only after the *event of interest* occurs

Estimators: parametric G-computation

Young et al, 2020

Under the classical assumptions of exchangeability, positivity, and consistency (and non informative censoring), the following parametric G-computation approach can be used by applying the following steps:

1. Fit a flexible parametric model (e.g. *pooled logistic)* that can be used to predict the conditional survival function given exposure, baseline covariates, and time to follow-up.
2. Predict the sub-distribution function at time *t* for each subject both under exposure and under no exposure, regardless of the subject's true exposure status.
3. Separately average the conditional sub-distribution functions under the exposure and under no exposure, over all subjects (*standardized sub-distribution function*)

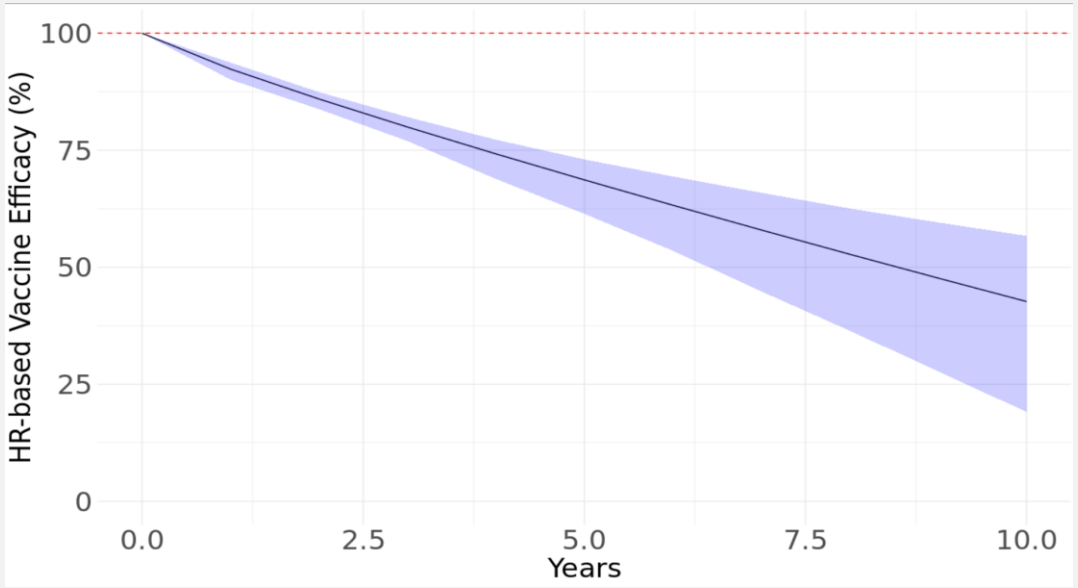
$$F_1(t \mid A = a) = \mathbb{E}[F_1(t \mid A = a, X)] = \frac{1}{N} \sum_{i=1}^{N} F_1(t \mid A = a, X = x_i)$$

It can be shown that the standardized sub-distribution hazard can be calculated as

$$h_1^{\text{sub}}(t \mid A = a) = \frac{\sum_{i=1}^{N} [1 - F_1(t \mid A = a, X = x_i)]\, h_1(t \mid A = a, X = x_i)}{N - \sum_{i=1}^{N} F_1(t \mid A = a, X = x_i)}$$

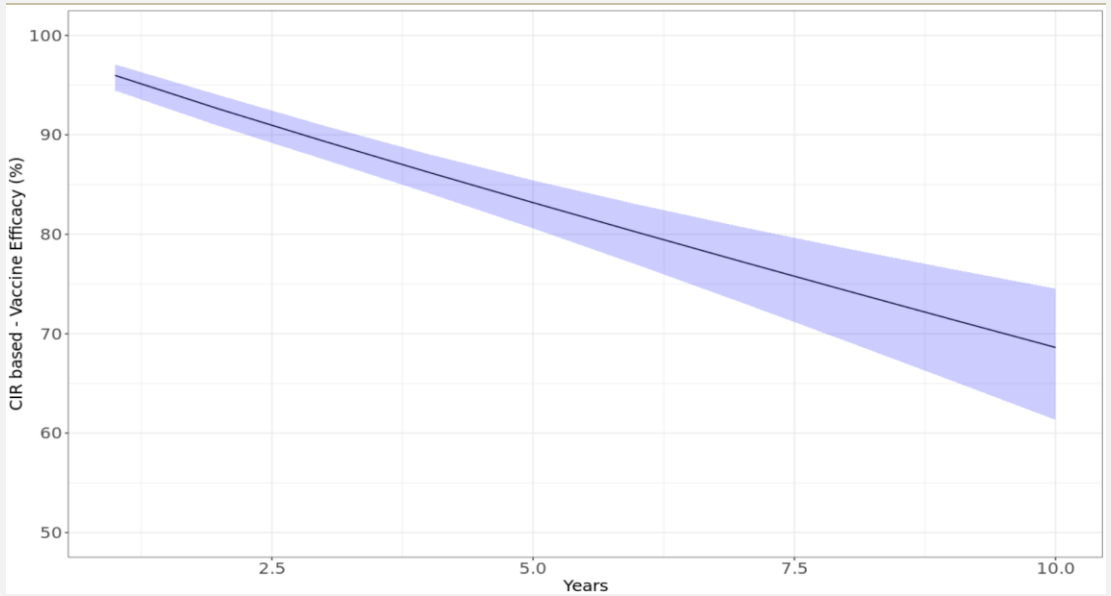
Result - Extrapolation results from different methodologies

Hazard Rate Ratio method



VE at time t for people who were vaccinated at time 0 and are still free of EOI at time t-1 - **45% 10 years following vaccination**

Cumulative Incidence Rate Ratio method

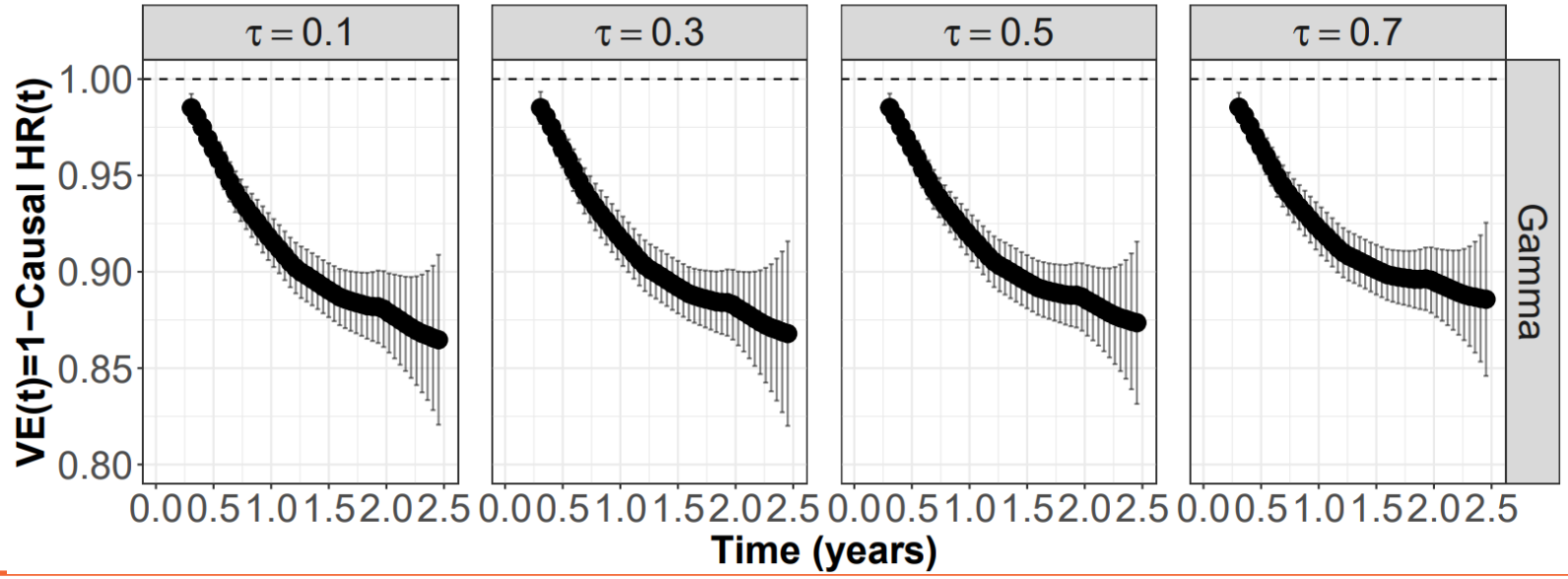


What is the overall VE over the duration of time 0 to time t - ***68% 10 years following vaccination***

Parametric model (flexible Weibull distribution selected based on minimum BIC)

Sensitivity analysis - Causal HR

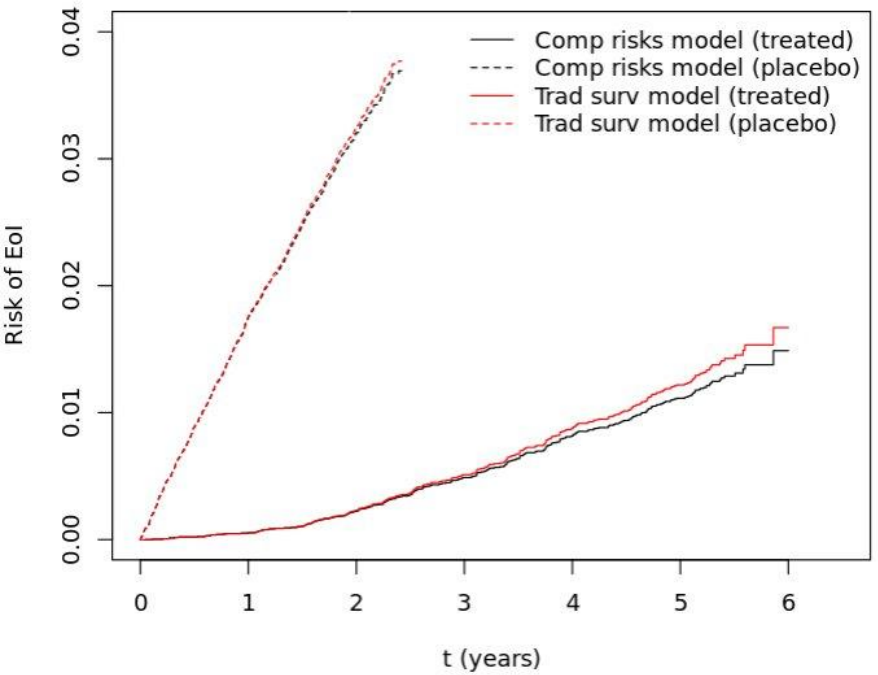
- **CHR is not identifiable, therefore, sensitivity analysis is needed.**
- Sensitivity analysis (R. Axelrod and D.Nevo, 2022*) has been performed using Phase III hypothetical data (no extension study)
- Estimator of causal effect has been obtained as a function of unidentified dependence between event times in two treatment groups (frailty model – gamma) with different values of Kendall Tau

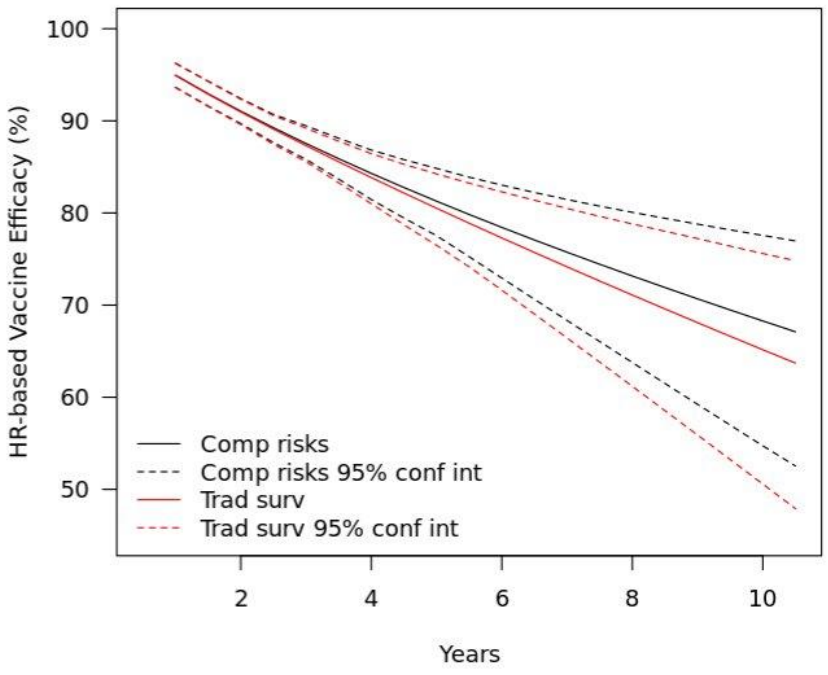


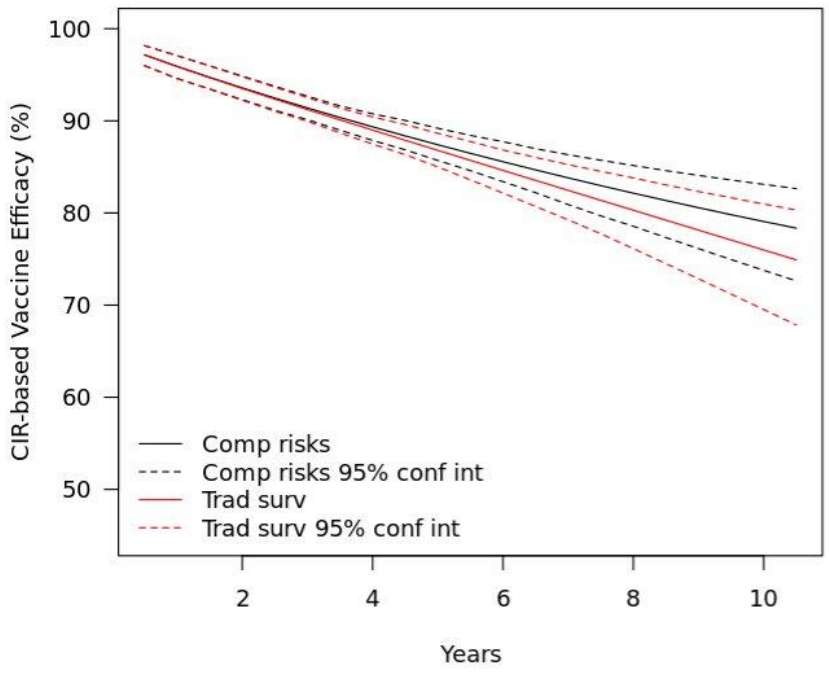
Analysis shows that with different values of Kendall Tau, the VE at each year doesn't change much suggesting minimal selection bias using the standard approach (which is conservative).

* A sensitivity analysis approach for the causal hazard ratio in randomized and observational studies

Estimated CIR and HR based Vaccine Efficacy considering competing risk







Parametric model (flexible pooled Logistic)

Slight increase in VE considering competing risk (Sub-distribution hazard ratio (HR) and Sub-distribution incidence ratio (CIR))

Conclusions

- In context of vaccine efficacy persistence study, we considered *flexible parametric G-computation* to estimate and extrapolate the VE
- Advantages of the proposed approach compared to the classical one
  - Estimated marginal effects (interpretability) adjusting for covariates (efficiency)
  - Data-driven shape of VE waning (flexible parametric model selection using BIC)
- Furthermore, we considered
  - Sensitivity analysis with respect to *causal HR* (R. Axelrod and D.Nevo, 2022)
  - Parametric G-computation competing risk analysis (Young et al, 2020)
- Robust extrapolation depends on maturity of follow-up

Proposed causal inference framework will play a key role on discussions with authorities about VE persistence and cost effectiveness analysis

Main References and Acknowledgement

- Axelrod, R., & Nevo, D. (2023). A sensitivity analysis approach for the causal hazard ratio in randomized and observational studies. *Biometrics*, 79(3), 2743-2756.
- Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology*, 21(1), 13-15.
- Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC
- Martinussen, T., Vansteelandt, S., & Andersen, P. K. (2020). Subtleties in the interpretation of hazard contrasts. *Lifetime Data Analysis*, 26, 833-855.
- Rutherford, Mark J, Paul C Lambert, Michael J Sweeting, Becky Pennington, Michael J Crowther, Keith R Abrams, and Nicholas R. Latimer. 2020. “NICE DSU Technical Support Document 21: Flexible Methods for Survival Analysis.”
- Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., & Hernán, M. A. (2020). A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in medicine*, 39(8), 1199-1236.

Acknowledgement: Alessia Eletti, Maria Costa, Wenji Pu, GSK Vaccines

Thank you