Covariate adjustment in time-to-event data, single and doubly-robust methods

Sanne Roels
Senior Principal Scientist
June 11th 2025
PSI Conference

Johnson&Johnson

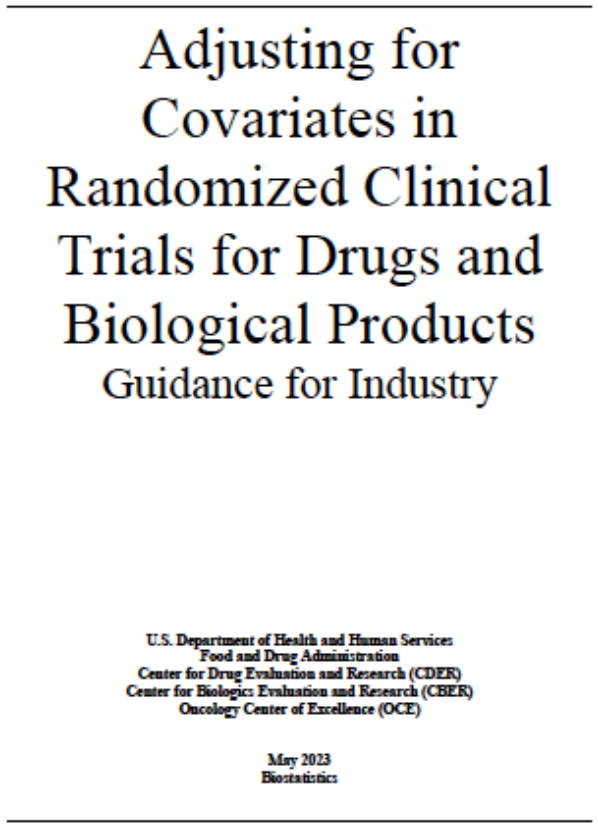
Disclaimer

Conflicts of interest: Sanne Roels is an employee of J&J and the views and opinions expressed in this presentation are solely those of the presenter and not necessarily those of J&J.

Acknowledgements

Thanks to Tina Sotto, Daniel Backenroth, Kelly Van Lancker.

Where to start?



Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

May 2023
Biostatistics

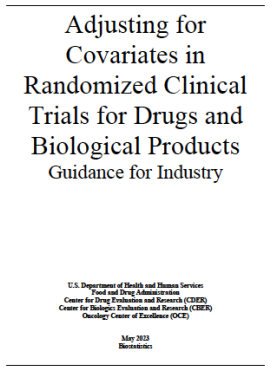
General Principles in the Guidance

1. Unadjusted remains valid …
2. Use models that are **robust against mis-specification**
3. Use **pre-specified** prognostic variables
4. Sponsors should discuss proposals for **complex** covariate-adaptive randomization, **data-adaptive covariate selection**, or use of covariate adjustment in an adaptive design with the relevant review division.

FDA guidance: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adjusting-covariates-randomized-clinical-trials-drugs-and-biological-products

What are the Options?

For Time-To-Event Outcomes


Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products
Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

May 2023
Biostatistics

- Covariate-adjusted estimators of unconditional treatment effects that are robust to misspecification of regression models have been proposed for randomized clinical trials with binary outcomes (e.g., Steingrimsson et al. 2017), ordinal outcomes (e.g., Díaz et al. 2016), count outcomes (e.g., Rosenblum and van der Laan 2010), and time-to-event outcomes (e.g., Tangen and Koch 1999; Lu and Tsiatis 2008). If a novel method is proposed and statistical properties are unclear, the specific proposal should be discussed with the review division.

What is in these papers?

How does it relate to the current practice and evolutions?

Covariate Adjusted LogRank Test

What is out there? (1)

- **Tangen & Koch (1999)**: Covariate adjusted LogRank and Wilcoxon test – ***testing*** only
  - Evaluation demonstrated its efficiency (Jiang et al, 2008).
- **Lu and Tsiatis (2008)**: Covariate adjusted LogRank and HR – **testing and estimation**
  - Equivalence partial likelihood test of Cox Score test (under PH) and Log-rank test
  - Implementation available in `speff2trial` package (CRAN)
- **Ye et al. (2024)** generalized this procedure to cover stratified design and achieve guaranteed efficiency gains
  - Implemented in the `RobinCar` package (on CRAN)

Tangen and Koch (1999) . Nonparametric analysis of covariance for hypothesis testing with logrank and Wilcoxon scores and survival-rate estimation in a randomized clinical trial, J. Biopharm Stat. doi:10.1081/BIP-100101179
Lu and Tsiatis (2008). Improving the Efficiency of the Log-Rank Test Using Auxiliary Covariates. Biometrika. https://www.jstor.org/stable/20441494
Ye et al. (2024) Covariate-adjusted log-rank test: guaranteed efficiency gain and universal applicability. doi.org/10.1093/biomet/asad045

Covariate Adjusted LogRank Test

What is out there? (2)

Advantages

- In line with current practice of using HR (Cox and LogRank)
- Straightforward implementation and pre-specification
- Software is available

Limitations

- Under *uninformative censoring (*and *proportional hazards* assumption for HR)
- Corresponding graphical display of survival curve?
- How many covariates versus risk of overfitting?

Survival Probabilities

Defined on the Counterfactual Survival Probabilities

- Survival probabilities are defined in a *counterfactual* way for both treatment arms
- Consider individual *i*, with corresponding survival time *T* for treatment arm A=*a* with baseline covariates.
- *The corresponding counterfactual survival probability evaluated at timepoint t for individual i with baseline covariates x_i :*
  - $S^a(t|x_i) = P(T^a > t\,|x_i)$,
  - with the counterfactual survival probability for population $E\big(I(T^a > t)\big)$
- *Population summary metric* can be expressed as an average treatment effect (additive, ratio, …)
- By calculating the the probabilities over a range of values, one can draw adjusted survival curves (e.g. Denz et al)

Denz et al. (2023). A comparison of different methods to adjust survival curves for confounders. Statistics in Medicine. doi: 10.1002/sim.9681

Alternatives to the Marginal HR

Efficiency Gains Observed

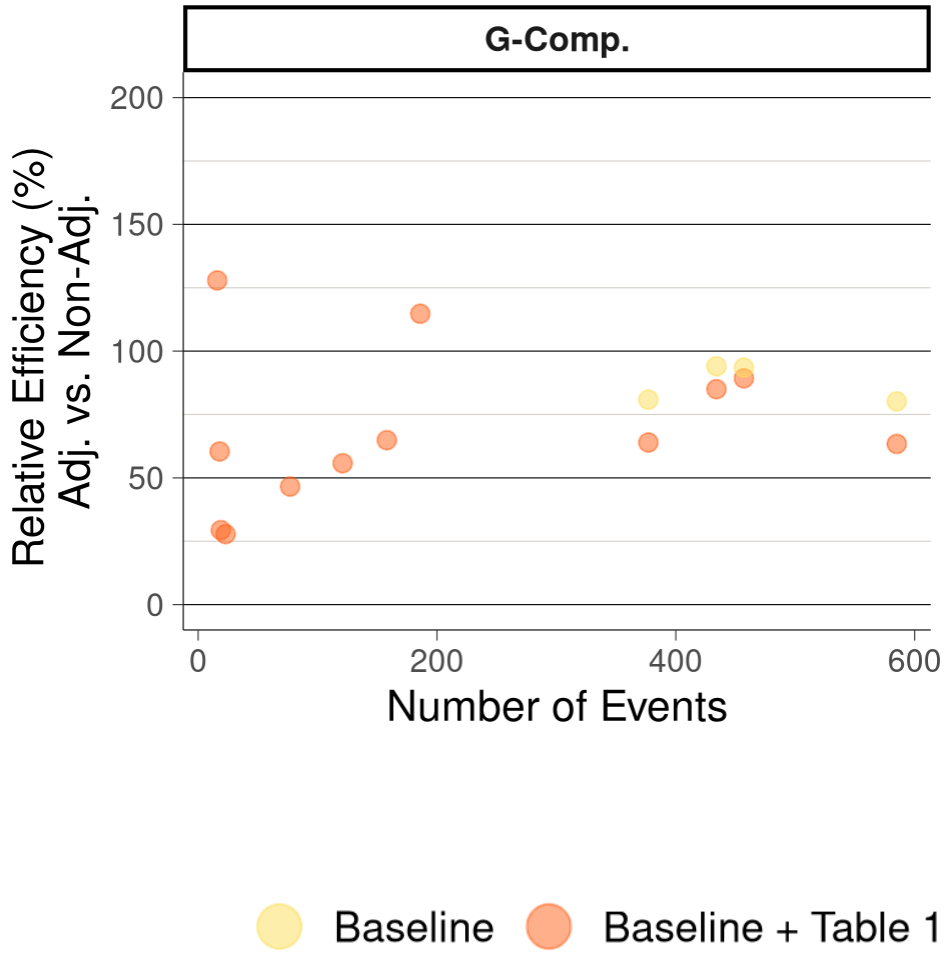
Efficiency = variance(adjusted estimate)/variance(unadjusted estimate)

Survival probabilities evaluated at the 75th quantile of the event time distribution

- One could consider ***G-computation*** for estimation
- Evaluation of efficiency adjusted analysis versus unadjusted analysis in 13 real datasets where a difference in survival probabilities was evaluated

- ***Brute force comparison with unadjusted analysis, in real studies with different variables considered***
  - ***Efficiency gains*** *are observed*

However, …



Alternatives to the Marginal HR

Contrasting Survival Probabilities on Specific Time point

- G-computation is not robust against misspecification (but, maybe reasonable in simple settings? See also Chen et al., 2023)
  - Weighting approaches (IPC/TW) are singly robust, but potentially less stable, impacting efficiency negatively

- An alternative single-robust methods exist: Iterative Conditional Expectation procedure
- Essentially, by breaking the time the survival into discrete time intervals and iteratively applying the g-formula (standardization) on these intervals, one obtains the event probability and survival probabilities (See e.g. Wen et al., 2020)

- In addition, doubly robust Targeted Maximum Likelihood Estimation approach exists too

Chen et al (2023). Beyond the Cox Hazard Ratio: A Targeted Learning Approach to Survival Analysis in a Cardiovascular Outcome Trial Application. Statistics in Biopharmaceutical Research. Doi:10.1080/19466315.2023.2173644
Wen et al. (2020). Parametric g-formula implementations for causal survival analyses. Biometrics. Doi: 10.1111/biom.13321

Alternatives to the HR

Add cox models

TMLE: Doubly Robust Estimation of Survival Probabilities

Early work: Moore and van der Laan (2009), Stittelman et al. (2011)

- Explicit focus on 3 elements in estimation: 1) shape of the survival curve; 2) *censoring* mechanism; 3) flexibility in the underlying models, allowing for ML
- Several TMLE implementations rely on time discretization, and then iteratively applies TMLE on a binary outcome in that interval, allowing maximal flexibility, and non-parametrically estimating the survival probability at specific time point
- Recently extended to continuous time space: Rytgaard et al, 2021)

Moore and van der Laan (2009), Increasing Power in Randomized Trials with Right-Censored Outcomes Through Covariate Adjustment. Journal of Biopharm. Statistics. doi:10.1080/10543400903243017
Stittelman et al. (2011). Targeted maximum likelihood estimation of effect modification parameters in survival analysis. International Journal of Biostatistics. Doi: 10.2202/1557-4679.1307
Rytgaard et al. (2022). Continuous-time targeted minimum loss-based estimation of intervention-specific mean outcomes. Annals of Statistics. DOI: 10.1214/21-AOS2114

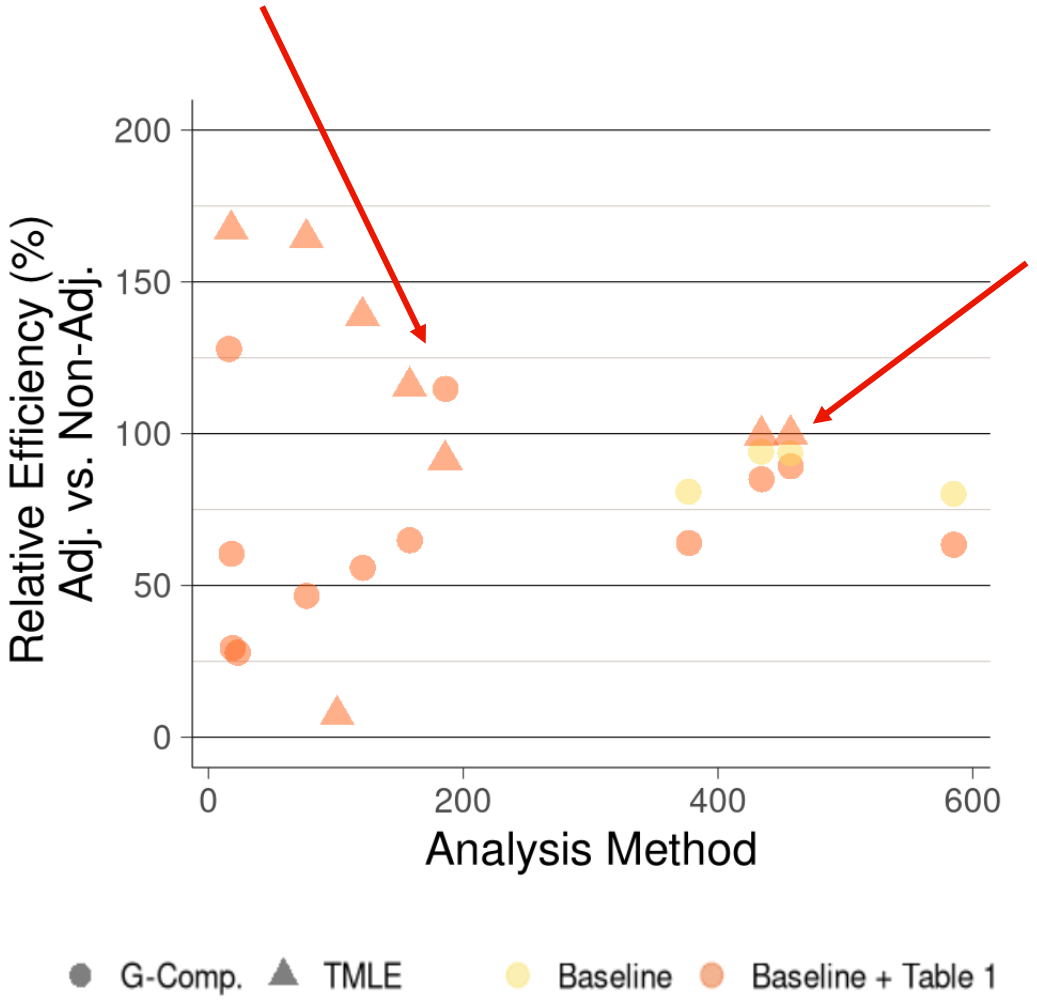
Alternatives to the HR

Performance of TMLE in clinical trials?

- Same brute force comparison,
- Some instances have substantial gain, other less so…

- Contrast vis-à-vis with G-computation; no clear pattern however

→ Note that the censoring assumptions are also different here (versus unadjusted KM) making a direct comparison not straightforward to interpret



Covariate adjusted Survival Probabilities

Summary

Advantages

- Easy interpretation of the summary metric
- Straightforward implementation and link with graphical display of the data
- Robustness is attainable with TMLE
- Software available

Limitations

- G-computation with the Cox model is not robust against misspecification
- How to avoid overfitting?
- Population summary metric is not regularly used in time-to-event settings
- TMLE has many procedural specifications, but pre-specification is possible (Gruber et al.,2023)

Gruber et al. (2022). Developing a Targeted Learning-Based Statistical Analysis Plan. Statistics in Biopharmaceutical Research. https://doi.org/10.1080/19466315.2022.2116104

Overview

Method	Testing	Summary Metric	Robust	Proportional Hazard	Censoring	Software
Covariate Adjusted Log Rank Test	V	none	V	X	Non-informative $C \perp (T, X) \mid A$	RobinCar, speff2trial
Covariate Adjusted HR	V	HR	V	V	Non-informative $C \perp (T, X) \mid A$	RobinCar, speff2trial
IPT/CW	V	HR / Based on $S^a(t)$	V	X	(Non-)informative $C \perp (T, X) \mid A$	Mets, AdjustedCurves
G-computation	V	Based on $S^a(t)$	X	X	Non-informative $C \perp (T, X) \mid A$	Mets, AdjustedCurves
TMLE based	V	Based on $S^a(t)$	V	X(/V)	(Non-) Informative	SurvTmleRct, Concrete

Where has this brought us?

Some Thoughts

Baseline covariates are routinely collected in RCT, so why not use these if they can increase efficiency, while maintaining a marginal estimand?

Across several estimation strategies efficiency gains are attained when covariates are included in the estimation/testing procedure

Covariate Adjusted logRank and HR have been compared with contrasts of survival probabilities, in non-robust and doubly robust settings

Note that assumptions and robustness properties differ between estimation methods, making a direct comparison not straightforward

Thank You!

sroels4@its.jnj.com

Johnson&Johnson