Assumption-lean covariate adjustment for time-to-event outcomes

Dominic Magirr

Basel Biometric Society Meeting 25th March 2025

NOVARTIS

Acknowledgements

Craig Wang

Alexander Przybylski

Xinlei Deng

Mark Baillie

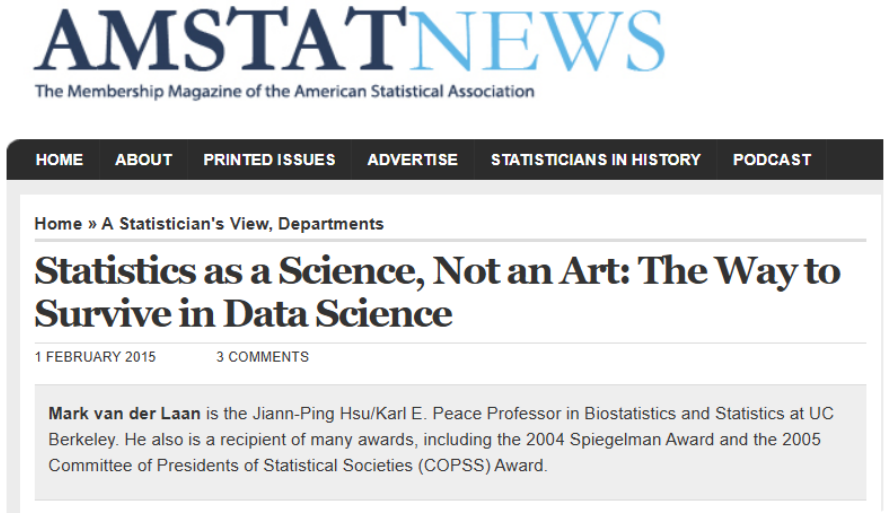
Tim Morris

Alex Ocampo

Ekkehard Glimm

NOVARTIS

The Science of Statistics


AMSTATNEWS

The Membership Magazine of the American Statistical Association

HOME ABOUT PRINTED ISSUES ADVERTISE STATISTICIANS IN HISTORY PODCAST

Home » A Statistician's View, Departments

Statistics as a Science, Not an Art: The Way to Survive in Data Science

1 FEBRUARY 2015 3 COMMENTS

Mark van der Laan is the Jiann-Ping Hsu/Karl E. Peace Professor in Biostatistics and Statistics at UC Berkeley. He also is a recipient of many awards, including the 2004 Spiegelman Award and the 2005 Committee of Presidents of Statistical Societies (COPSS) Award.

https://magazine.amstat.org/blog/2015/02/01/statscience_feb2015/

The foundation of statistics laid down by its founders [...] could not have been to arbitrarily select a "convenient" statistical model. However, that is precisely what most statisticians blithely do, proudly referring to the quote, "All models are wrong, but some are useful."

[...]

one typically asks a few questions about the data such as: Is the outcome a survival time? Is it case-control data? And then one quickly moves on to returning output from a Cox-Ph model or a logistic regression model with some "reasonable" set of covariates

[...]

Is this mess we have created really necessary? No! As a start, we need to take the field of statistics (i.e., the science of learning from data) seriously. It is complete nonsense to state that all models are wrong, so let's stop using that quote. For example, a statistical model that makes no assumptions is always true.

Roadmap for this talk

Status quo

(Stratified) Cox models to estimate conditional hazard ratios

FDA Position

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

"Model-free estimand" and "assumption-lean analysis"

Compare unconditional probability of survival (or restricted mean survival time) on the two treatment arms.

Double-robust covariate-adjusted estimators: AIPCW, TMLE

What's in the guidance?

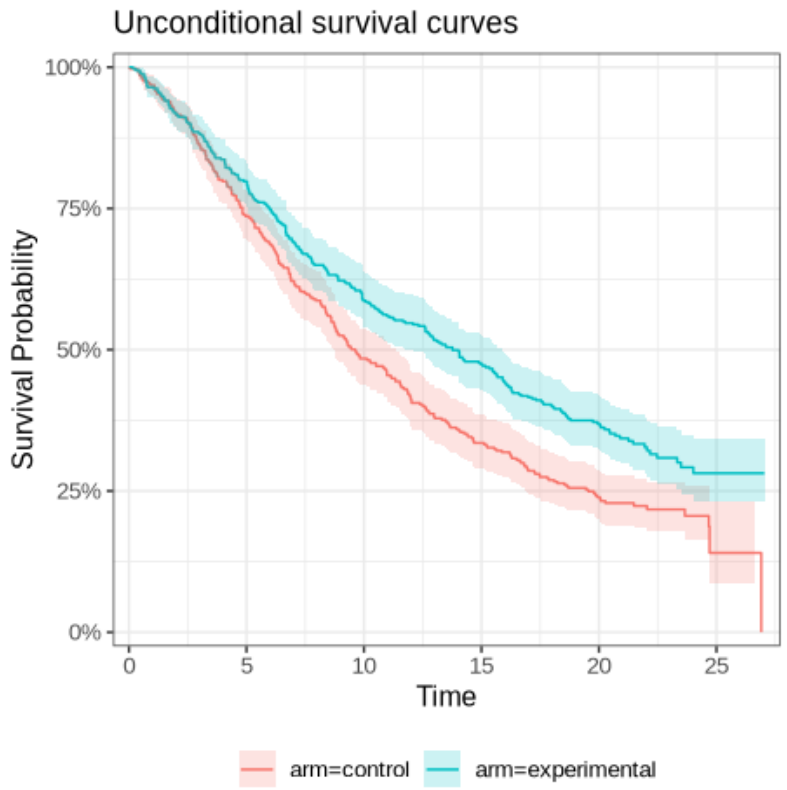
How to implement?

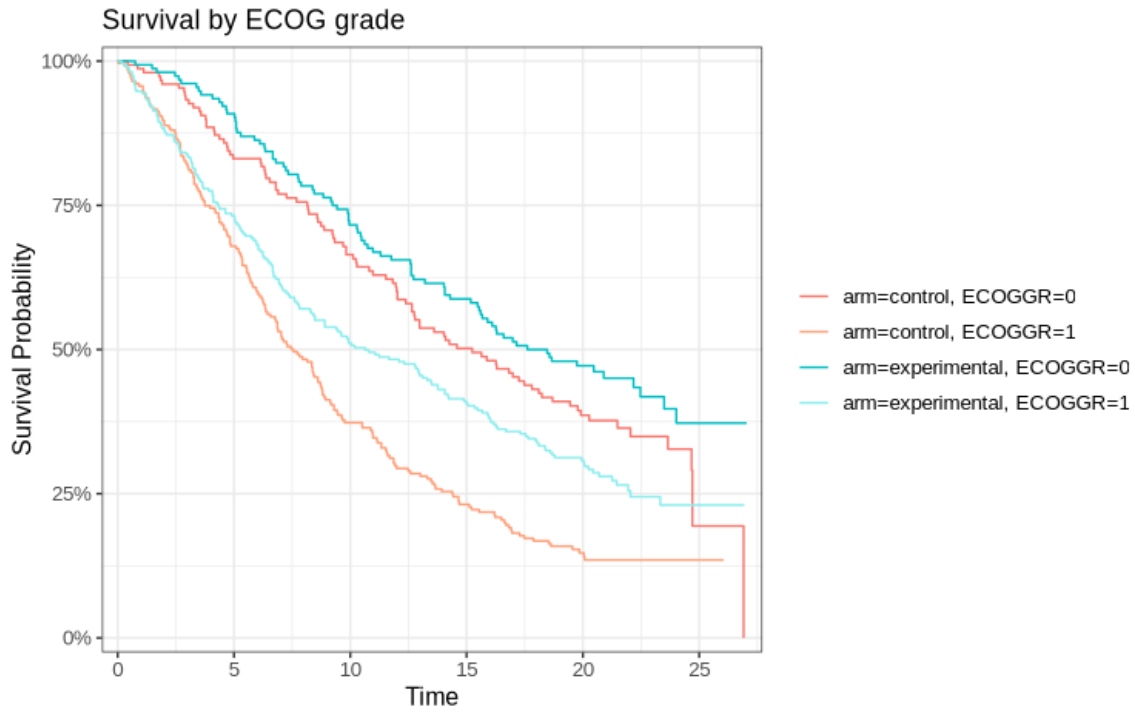
Benefits / challenges?

Should we go further?

Status quo: example based on the OAK trial

"The HR was estimated with a stratified Cox regression analysis. Stratification factors were the same used for randomisation."

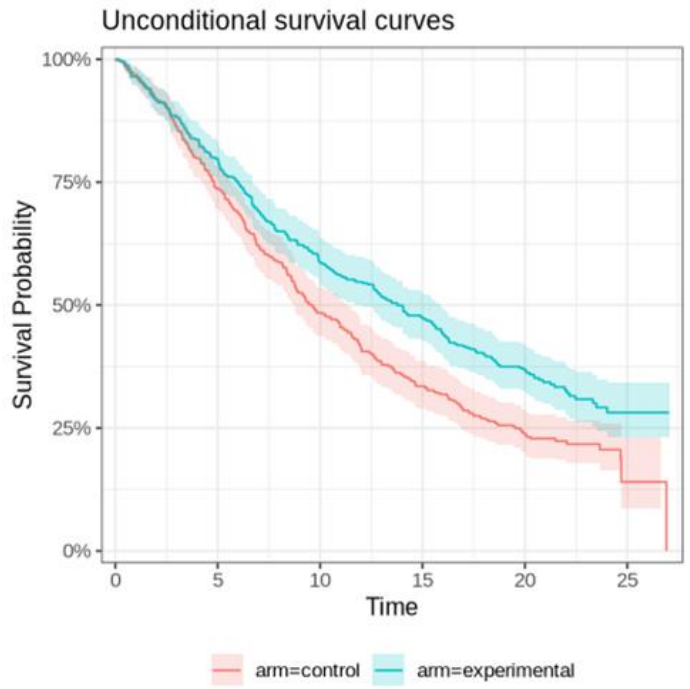


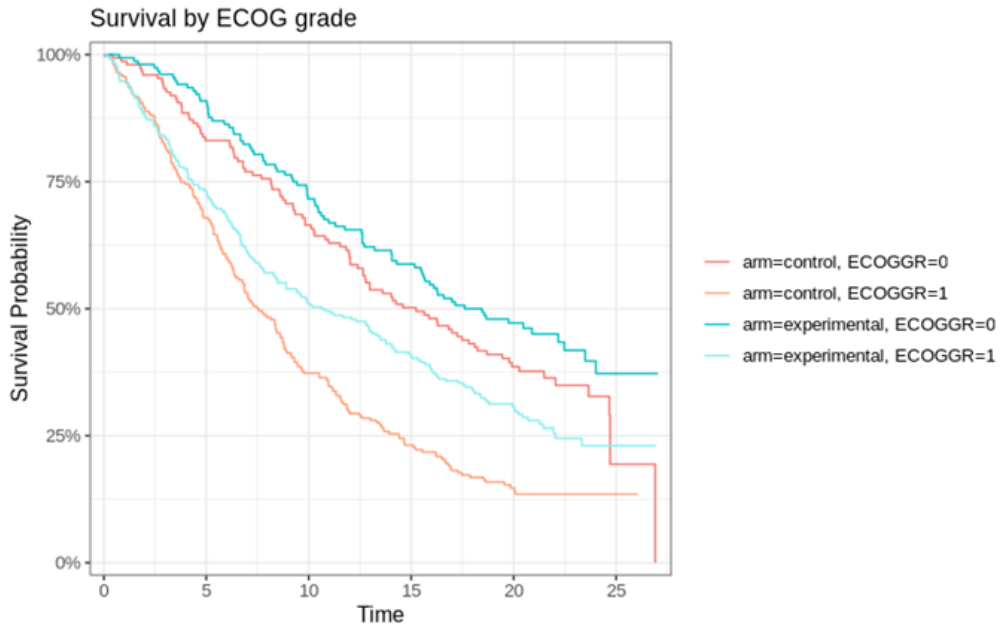


Rittmeyer et al. 2017 https://doi.org/10.1016/S0140-6736(16)32517-X; Gandara et al. 2018 https://doi.org/10.1038/s41591-018-0134-3

NOVARTIS

Status quo: example based on the OAK trial





$$h(t) = h_0(t)\exp(\beta \times \mathbb{I}(\mathbf{trt}))$$

$$\exp(\hat{\beta}) = 0.729;\ \mathrm{se}(\hat{\beta}) = 0.0842;\ Z = -3.76$$

$$h(t) = \{\mathrm{ECOG} \times h_{0,1}(t) + (1 - \mathrm{ECOG}) \times h_{0,0}(t)\}\exp(\beta^{(S)} \times \mathbb{I}(\mathbf{trt}))$$

$$\exp(\widehat{\beta^{(S)}}) = 0.713;\ \mathrm{se}(\widehat{\beta^{(S)}}) = 0.0845;\ Z = -4.00$$

NOVARTIS

Health authority guidelines on covariate adjustment

1998

INTERNATIONAL CONFERENCE ON HARMONISATION OF TECHNICAL REQUIREMENTS FOR REGISTRATION OF PHARMACEUTICALS FOR HUMAN USE

ICH HARMONISED TRIPARTITE GUIDELINE

STATISTICAL PRINCIPLES FOR CLINICAL TRIALS
E9

2015

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

26 February 2015
EMA/CHMP/295050/2013
Committee for Medicinal Products for Human Use (CHMP)

Guideline on adjustment for baseline covariates in clinical trials

2023

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

2020

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

17 February 2020
EMA/CHMP/ICH/436221/2017
Committee for Medicinal Products for Human Use

ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials

Key point (arguably) from FDA 2023 guidance

Emphasis on unconditional estimands

- Sponsors can perform covariate-adjusted estimation and inference for an unconditional treatment effect (e.g., the odds ratio of 4.8 in Table 1) in the primary analysis of data from a randomized trial.

- Sponsors should discuss with the relevant review divisions specific proposals in a protocol or statistical analysis plan containing nonlinear regression to estimate conditional treatment effects for the primary analysis.

Table 1: Non-collapsibility of the Odds Ratio in a Hypothetical Target Population

	Percentage of target population	Success rate		Odds ratio
		New drug	Placebo	
Biomarker-positive	50%	80.0%	33.3%	8.0
Biomarker-negative	50%	25.0%	4.0%	8.0
Combined	100%	52.5%	18.7%	4.8

"When estimating a conditional treatment effect through nonlinear regression, the model assumptions will generally not be exactly correct, and results can be difficult to interpret if the model is misspecified"

Model-trusting

$$\text{logit}\, P(Y = 1 \mid A, X) = \alpha_0 + \alpha_1 A + \alpha_2 X$$

- **Direct** estimation via MLE / posterior probability
- If model is incorrect, it's unclear what α_1 means
- Compatible with Bayesian, likelihood, and frequentist (conditional and unconditional) inference
- Typically used for conditional estimands

Model-robust / assumption-lean

$$\overline{Y}_1 - \overline{Y}_0 + \frac{1}{n}\sum_{i=1}^{n}\left(A_i - \frac{1}{2}\right)h(X_i)$$

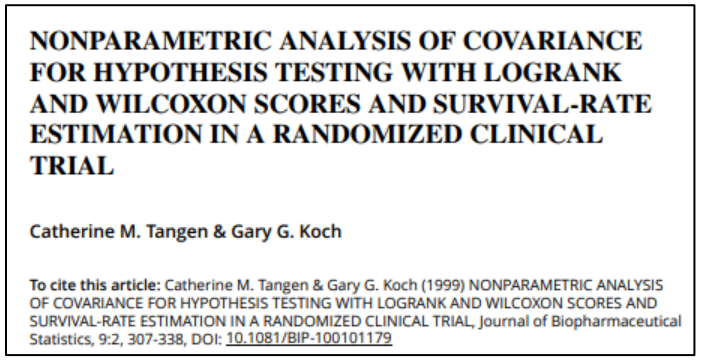
- Combines an unadjusted estimator with an "estimator of zero"
- Clever choice of h(x) to increase efficiency
- An (unconditional) frequentist approach
- Typically used for unconditional estimands

Buja et al. (2019); Vansteelandt (2021)

What about time-to-event outcomes?

FDA guidance

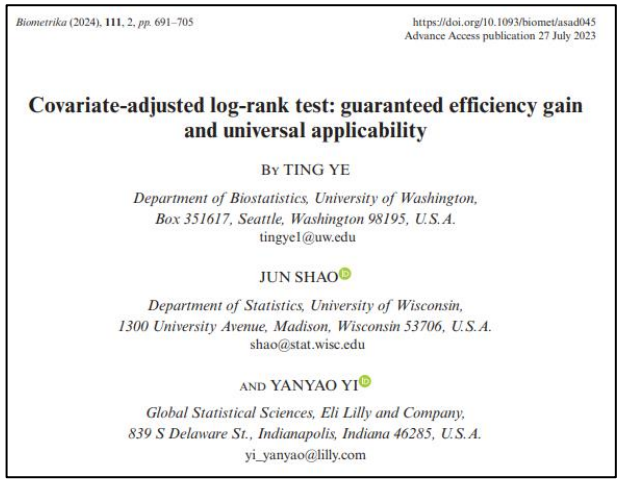
- Covariate-adjusted estimators of unconditional treatment effects that are robust to misspecification of regression models have been proposed for randomized clinical trials with binary outcomes (e.g., Steingrimsson et al. 2017), ordinal outcomes (e.g., Díaz et al. 2016), count outcomes (e.g., Rosenblum 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.



NONPARAMETRIC ANALYSIS OF COVARIANCE FOR HYPOTHESIS TESTING WITH LOGRANK AND WILCOXON SCORES AND SURVIVAL-RATE ESTIMATION IN A RANDOMIZED CLINICAL TRIAL

Catherine M. Tangen & Gary G. Koch

To cite this article: Catherine M. Tangen & Gary G. Koch (1999) NONPARAMETRIC ANALYSIS OF COVARIANCE FOR HYPOTHESIS TESTING WITH LOGRANK AND WILCOXON SCORES AND SURVIVAL-RATE ESTIMATION IN A RANDOMIZED CLINICAL TRIAL, Journal of Biopharmaceutical Statistics, 9:2, 307-338, DOI: 10.1081/BIP-100101179



Biometrika (2024), **111**, 2, *pp.* 691–705 https://doi.org/10.1093/biomet/asad045
Advance Access publication 27 July 2023

Covariate-adjusted log-rank test: guaranteed efficiency gain and universal applicability

BY TING YE

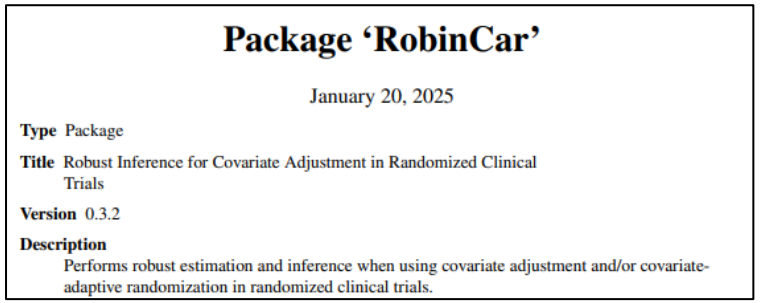
Department of Biostatistics, University of Washington, Box 351617, Seattle, Washington 98195, U.S.A.
tingye1@uw.edu

JUN SHAO

Department of Statistics, University of Wisconsin, 1300 University Avenue, Madison, Wisconsin 53706, U.S.A.
shao@stat.wisc.edu

AND YANYAO YI

Global Statistical Sciences, Eli Lilly and Company, 839 S Delaware St., Indianapolis, Indiana 46285, U.S.A.
yi_yanyao@lilly.com



Package 'RobinCar'

January 20, 2025

Type Package

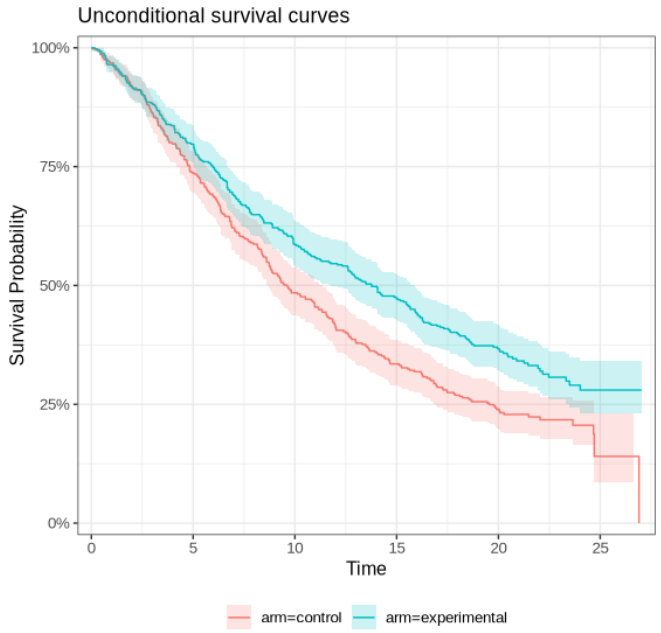
Title Robust Inference for Covariate Adjustment in Randomized Clinical Trials

Version 0.3.2

Description
Performs robust estimation and inference when using covariate adjustment and/or covariate-adaptive randomization in randomized clinical trials.

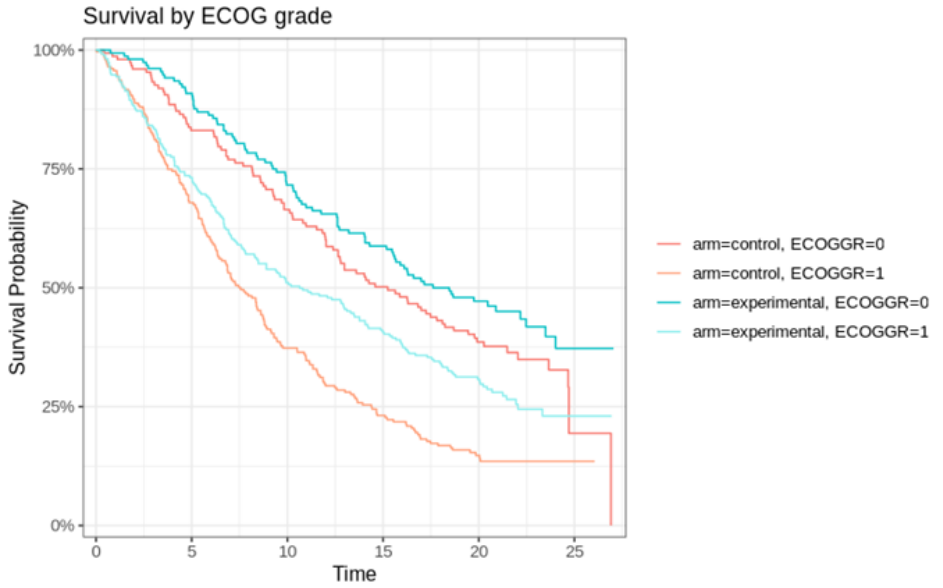
https://marlenabannick.com/RobinCar/index.html

Covariate-adjusted estimator of unconditional hazard ratio



$$h(t) = h_0(t)\exp(\beta \times \mathbb{I}(\mathbf{trt}))$$

$$\exp(\hat{\beta}) = 0.729;\ \mathrm{se}(\hat{\beta}) = 0.0842;\ Z = -3.76$$

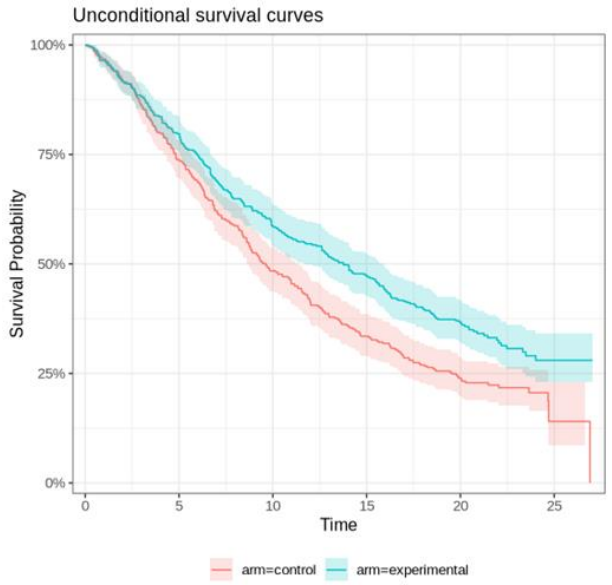


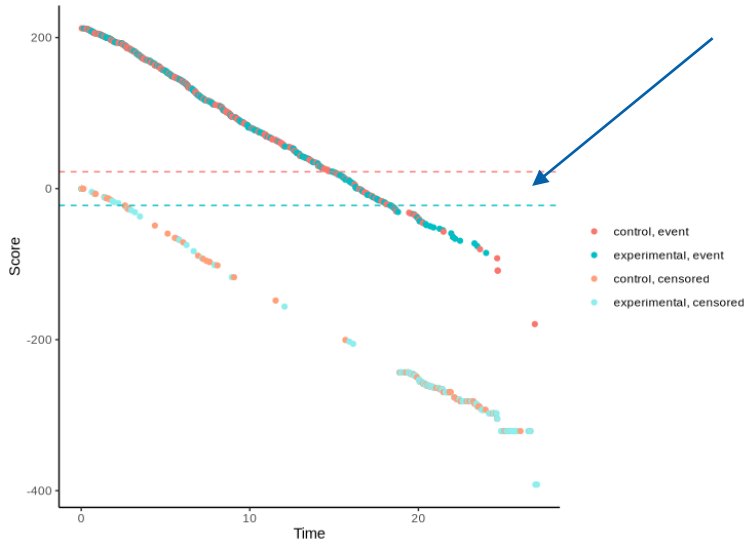
$$h(t) = h_0(t)\exp(\beta \times \mathbb{I}(\mathbf{trt}))$$

$$\exp(\hat{\beta}_{adj}) = 0.724;\ \mathrm{se}(\hat{\beta}_{adj}) = 0.0816;\ Z = -3.97$$

How does this work?

Hypothesis testing





Difference in average score is LR($\beta = 0$)

$$h(t) = h_0(t)\exp(\beta \times \mathbb{I}(\mathbf{trt}))$$

$$LR(\beta = 0) = \sum_{t_j} O_{1,j} - E_{1,j} = \cdots = \frac{\sum_{i=1}^{n} A_i Score_i}{\sum_{i=1}^{n} A_i} - \frac{\sum_{i=1}^{n}(1 - A_i)Score_i}{\sum_{i=1}^{n}(1 - A_i)}$$

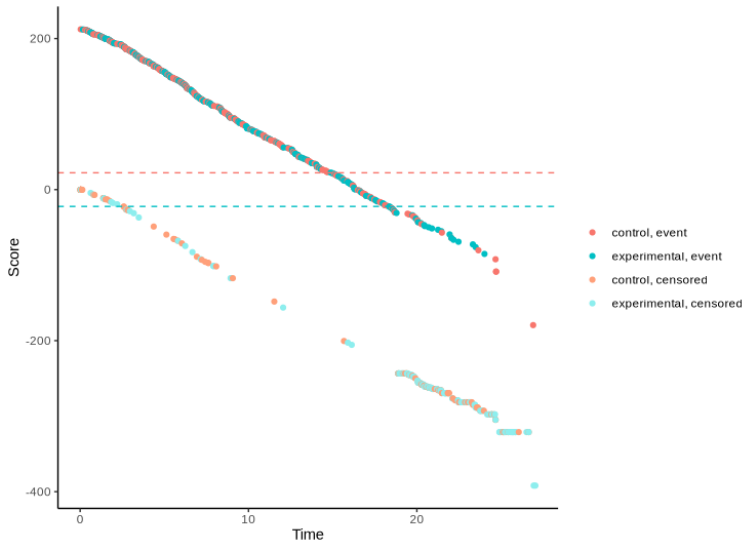
Expected #events on trt 1 under H_0: $\beta = 0$

See, e.g., Leton & Zuluaga (2001)

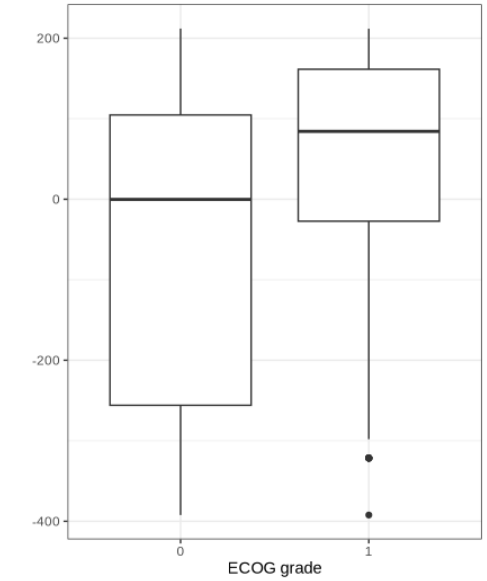
NOVARTIS

How does this work?

Hypothesis testing



Can baseline covariates explain any of the variation in scores?



Suggests a linear model: $\mathrm{E}(Score_i|A_i, ECOG_i) = \gamma + \gamma_1 A_i + \gamma_2 A_i(ECOG_i - \overline{ECOG}) + \gamma_3(1 - A_i)(ECOG_i - \overline{ECOG})$

If $|\widehat{\gamma_1}|$ large enough, reject H_0: $\beta = 0$

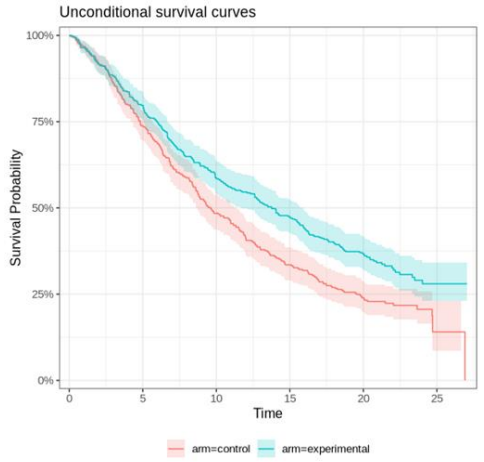
$$\widehat{\gamma_1} = LR - \{\hat{\gamma}_2(\overline{ECOG}_{trt=1} - \overline{ECOG}) - \hat{\gamma}_3(\overline{ECOG}_{trt=0} - \overline{ECOG})\}$$

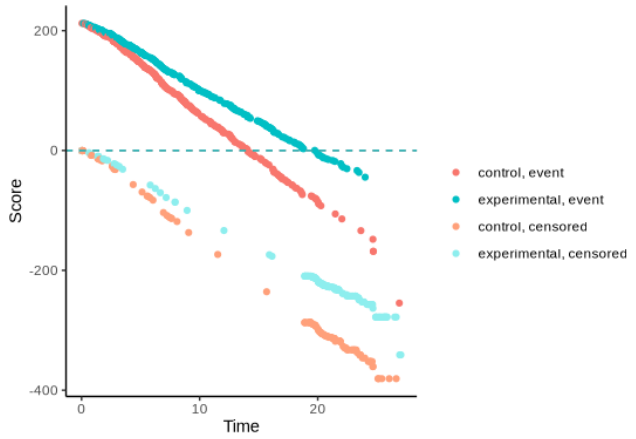
NOVARTIS

How does this work?

Point estimation

$$h(t) = h_0(t)\exp(\beta \times \mathbb{I}(\mathbf{trt}))$$





$$LR(\beta = \beta^*) = \sum_{t_j} O_{1,j} - E_{1,j}(\beta^*) = \cdots = \frac{\sum_{i=1}^{n} A_i Score_i(\beta^*)}{\sum_{i=1}^{n} A_i} - \frac{\sum_{i=1}^{n}(1-A_i)Score_i(\beta^*)}{\sum_{i=1}^{n}(1-A_i)}$$

Unadjusted: find β^* such that

$$LR(\beta^*) = 0$$

Adjusted: find β^* such that:

$$LR(\beta^*) - \{\hat{\gamma}_2(\overline{ECOG}_{trt=1} - \overline{ECOG}) - \hat{\gamma}_3(\overline{ECOG}_{trt=0} - \overline{ECOG})\} = 0$$

NOVARTIS

Implementation

- Ye et al. (2024): refinement of Tang & Koch (1999) + asymptotic theory
- Bannick et al. (2024): implementation in {RobinCar}

```
RobinCar::robincar_covhr(dat,
                         treat_col = "arm",
                         response_col = "time",
                         event_col = "event",
                         covariate_cols = NULL)

RobinCar::robincar_covhr(dat,
                         treat_col = "arm",
                         response_col = "time",
                         event_col = "event",
                         covariate_cols = "ECOGGR")

RobinCar::robincar_covhr(dat,
                         treat_col = "arm",
                         response_col = "time",
                         event_col = "event",
                         covariate_cols = c("ECOGGR", "blSLD"))
```

$$\exp(\hat{\beta}) = 0.729;\ \mathrm{se}(\hat{\beta}) = 0.0842;\ Z = -3.76$$

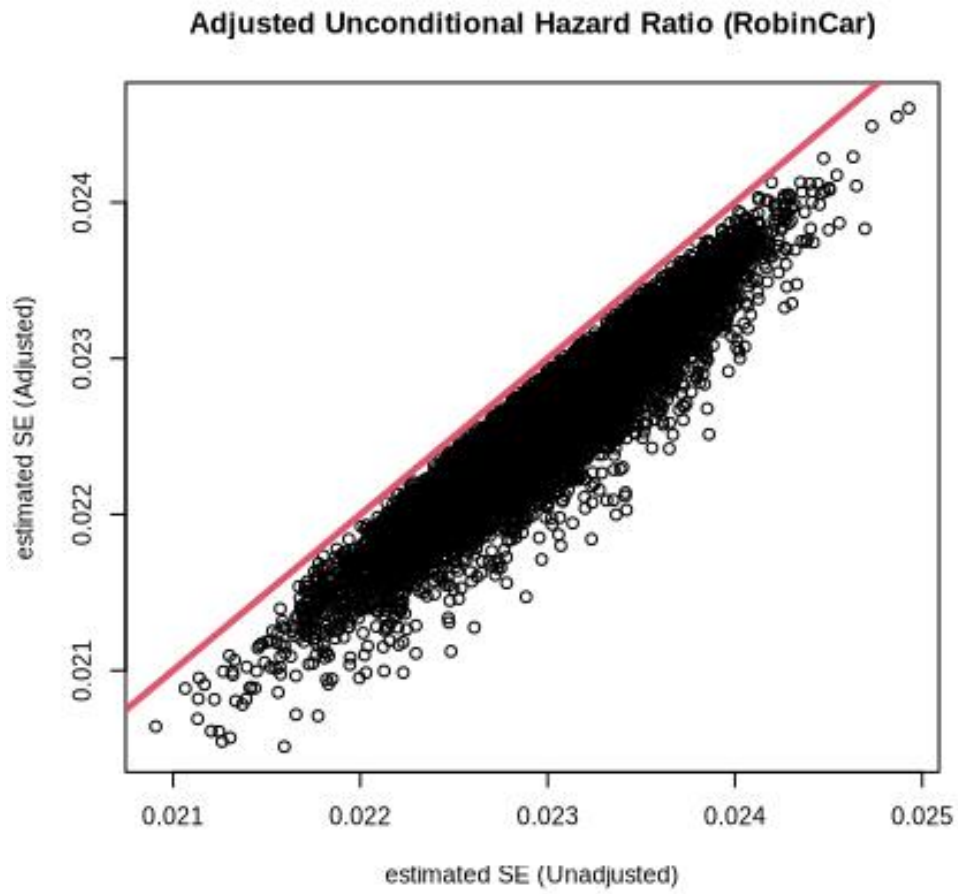
$$\exp(\hat{\beta}_{adj}) = 0.724;\ \mathrm{se}(\hat{\beta}_{adj}) = 0.0816;\ Z = -3.97$$

$$\exp(\hat{\beta}_{adj}) = 0.725;\ \mathrm{se}(\hat{\beta}_{adj}) = 0.079;\ Z = -4.07$$

Key opportunity: adjust for continuous covariates such as baseline tumour size (or also supercovariates) without changing the target estimand.

Caution: use responsibly in smaller sample sizes

- 10000 simulated trials
- Sample size 300; events ≈ 180; HR = 1
- Adjust for 10 covariates, all simulated from N(0, 1)



What about the proportional hazards assumption?


JOURNAL ARTICLE ACCEPTED MANUSCRIPT

Invited Commentary: Why use methods that require proportional hazards?

Mats J Stensrud ✉, Miguel A Hernàn

American Journal of Epidemiology, kwae361, https://doi.org/10.1093/aje/kwae361


Beyond the Cox Hazard Ratio: A Targeted Learning Approach to Survival Analysis in a Cardiovascular Outcome Trial Application

David Chen[a], Maya L. Petersen[a], Helene Charlotte Rytgaard[b], Randi Grøn[c], Theis Lange[b], Søren Rasmussen[c], Richard E. Pratley[d], Steven P. Marso[e], Kajsa Kvist[c], John Buse[f], and Mark J. van der Laan[a]

Statistics in Biopharmaceutical Research 15 (3): 524–39

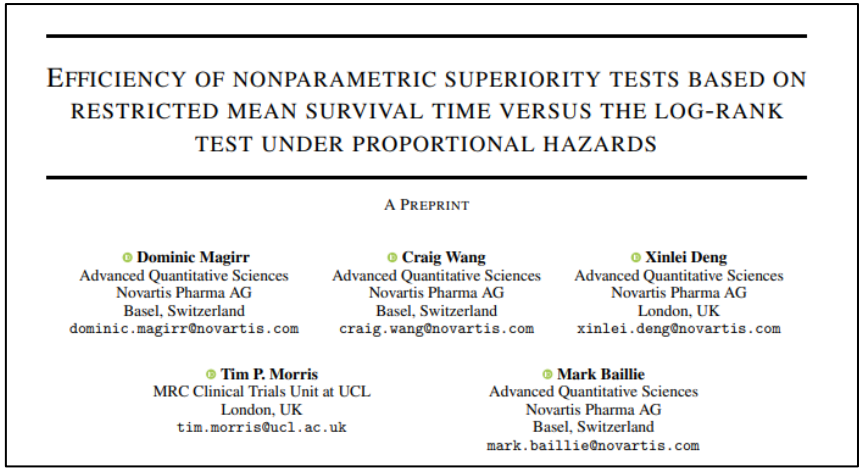
1) *PH assumption is not reasonable so why consider it?*

2) *PH assumption is not needed so why make it?*

See LinkedIn post by Stephen Senn for further discussion.

- *continued adherence to the Cox HR as a survival estimand is becoming increasingly indefensible.*
- *In simulations we demonstrated the double robustness and efficiency properties of TMLE.*
- *Our parallel reanalysis of LEADER trial data then reassuringly demonstrated that a TMLE targeting relative risk when compared to Cox provides compatible but more precise estimates of treatment effects,* ***even in a setting where Cox is expected to perform well.***

How efficient are model-free, assumption-lean methods?



EFFICIENCY OF NONPARAMETRIC SUPERIORITY TESTS BASED ON RESTRICTED MEAN SURVIVAL TIME VERSUS THE LOG-RANK TEST UNDER PROPORTIONAL HAZARDS

A PREPRINT

Dominic Magirr
Advanced Quantitative Sciences
Novartis Pharma AG
Basel, Switzerland
dominic.magirr@novartis.com

Craig Wang
Advanced Quantitative Sciences
Novartis Pharma AG
Basel, Switzerland
craig.wang@novartis.com

Xinlei Deng
Advanced Quantitative Sciences
Novartis Pharma AG
London, UK
xinlei.deng@novartis.com

Tim P. Morris
MRC Clinical Trials Unit at UCL
London, UK
tim.morris@ucl.ac.uk

Mark Baillie
Advanced Quantitative Sciences
Novartis Pharma AG
Basel, Switzerland
mark.baillie@novartis.com

https://arxiv.org/pdf/2412.06442

- Just a starting point: no covariates
- Contrary to some previous claims, there are realistic RCT settings where RMST-based methods do lose efficiency under the PH assumption.
- Next step: do these findings extend to covariate-adjusted estimators?

Summary

- **FDA guideline is an invitation to include continuous covariates (or supercovariates) into the primary analysis of RCTs with time-to-event endpoints. We should take advantage of this opportunity.**
  - Increases power (relative to unadjusted HR analysis).
  - Fits into established ways of designing RCTs.
  - Software implementation in {RobinCar}.
  - Use responsibly with a small number of covariates – what we think is most prognostic.
- **This is not fully aligned with the trend towards "model-free estimands, assumption-lean analysis" methods.**
  - Would be a more radical change in study design and analysis.
  - Involves trade-offs in power under different scenarios – this should be made transparent.
  - Software implementation currently lacking.
  - Let's get prepared to be able to analyse trials this way: requires investment in teaching and software.

References

Ozenne, B. M. H., Scheike, T. H., Stærk, L., & Gerds, T. A. (2020). On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biometrical Journal*, *62*(3), 751-763.

Ye, T., Shao, J., & Yi, Y. (2024). Covariate-adjusted log-rank test: guaranteed efficiency gain and universal applicability. *Biometrika*, *111*(2), 691-705.

Stensrud, M. J., & Hernàn, M. A. (2025). Invited Commentary: Why use methods that require proportional hazards?. *American Journal of Epidemiology*, kwae361.

Chen, D., Petersen, M. L., Rytgaard, H. C., Grøn, R., Lange, T., Rasmussen, S., ... & van der Laan, M. J. (2023). Beyond the Cox Hazard Ratio: A Targeted Learning Approach to Survival Analysis in a Cardiovascular Outcome Trial Application. *Statistics in Biopharmaceutical Research*, *15*(3), 524-539.

Tangen, C. M., & Koch, G. G. (1999). Nonparametric analysis of covariance for hypothesis testing with logrank and Wilcoxon scores and survival-rate estimation in a randomized clinical trial. *Journal of Biopharmaceutical Statistics*, *9*(2), 307-338.

Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., Von Pawel, J., ... & Gandara, D. R. (2017). Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet*, *389*(10066), 255-265.

Gandara, D. R., Paul, S. M., Kowanetz, M., Schleifman, E., Zou, W., Li, Y., ... & Shames, D. S. (2018). Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nature medicine*, *24*(9), 1441-1448.

Tsiatis, A. A., & Davidian, M. (2007). Comment: Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *22*(4), 569.

Buja, A., Brown, L., Berk, R., George, E., Pitkin, E., Traskin, M., ... & Zhao, L. (2019). Models as approximations I. *Statistical Science*, *34*(4), 523-544.

Vansteelandt, S. (2021). Statistical modelling in the age of data science. *Observational Studies*, *7*(1), 217-228.

Leton, E., & Zuluaga, P. (2001). Equivalence between score and weighted tests for survival curves. *Communications in Statistics-Theory and Methods*, *30*(4), 591-608.

Bannick M, Ye T, Yi Y, Bian F (2024). *RobinCar: ROBust INference for Covariate Adjustment in Randomized clinical trials*. R package version 0.3.0.

Double robustness: idea

Tsiatis & Davidian, 2007

Estimand $\boldsymbol{S_1(12)}$; let $Y_i(1) = I(T_i(1) > 12)$. If we know assignment model $\boldsymbol{\pi_1(x) = P(A = 1|X = x)}$ then:

$$\hat{S}_1(12) = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i}{\pi_1(X_i)} Y_i + \frac{1}{n}\sum_{i=1}^{n} (A_i - \pi_1(X_i))h(X_i) \quad (1)$$

are all consistent. If we know the outcome model $\boldsymbol{m_1(x) = E(I(T \geq 12)|X = x, A = 1)}$ then:

$$\hat{S}_1(12) = \frac{1}{n}\sum_{i=1}^{n} m_1(X_i) + \frac{1}{n}\sum_{i=1}^{n} A_i(Y_i - m_1(X_i))g(X_i) \quad (2)$$

are all consistent. If we want estimator in form (1) **and** (2) then the **only** choice is $g(x) = \frac{1}{\pi_1(x)}$ which makes $h(x) = \frac{-m_1(x)}{\pi_1(x)}$.

Right censoring adds complexity but conceptually the same (see e.g., Ozenne 2020)