# Ensuring valid inference for Cox hazard ratios after variable selection

UNIVERSITEIT GENT

Kelly Van Lancker
Joint work with
Stijn Vansteelandt and Oliver Dukes

FACULTY OF SCIENCES

# My perspective on estimands, identification and estimation

| Estimand | Identification | Estimation |
| --- | --- | --- |
| Model-free | Plausible required assumptions | Data-adaptive<br>"Let data speak"<br>Gain efficiency |

# Outline

# Conditional versus unconditional estimands

| | Unconditional | Conditional |
|---|---|---|
| Estimand | Model-free<br>Single number | Often model-based<br>Surface |
| Interpretation | Simple (?) interpretation. | More comprehensive understanding of individual treatment effect |
| Drug approval decisions | Only if target population is similar to the RCT population. | More transportable. |

# Randomized trials with time-to-event endpoints

- Logrank test is gold standard.

# Randomized trials with time-to-event endpoints

- Logrank test is gold standard.
- Cox proportional hazards model also often used, allows to adjust for covariates (denoted by $L$).

# Randomized trials with time-to-event endpoints

- Logrank test is gold standard.
- Cox proportional hazards model also often used, allows to adjust for covariates (denoted by $L$).
  - Usually pre-specified.

**Censoring**

Changing the adjustment set changes our censoring assumption!

- Logrank test: censoring is (statistically) independent of survival time
- Cox model adjusting for $L$: censoring is independent of survival time, given treatment $A$ and baseline covariate $L$

# Adjustment set matters

## Censoring

Changing the adjustment set changes our censoring assumption!

- In practice, Cox model is often pre-specified.

# Adjustment set matters

**Censoring**

Changing the adjustment set changes our censoring assumption!

- In practice, Cox model is often pre-specified.
- This raises concerns about:

# Adjustment set matters

## Censoring

Changing the adjustment set changes our censoring assumption!

- In practice, Cox model is often pre-specified.
- This raises concerns about:
  - Censoring assumption: Can we assume non-informative censoring conditional on the variables in our model?

# Adjustment set matters

## Censoring

Changing the adjustment set changes our censoring assumption!

- In practice, Cox model is often pre-specified.
- This raises concerns about:
  - Censoring assumption: Can we assume non-informative censoring conditional on the variables in our model?
  - Model misspecification

# Adjustment set matters

**Censoring**

Changing the adjustment set changes our censoring assumption!

- In practice, Cox model is often pre-specified.
- This raises concerns about:
  - Censoring assumption: Can we assume non-informative censoring conditional on the variables in our model?
  - Model misspecification
- Variable selection procedures can help in choosing a model (with the right variables)!

# Data-adaptive methods

- Consider the hazard function for the Cox PH model

$$\lambda\{t|A, L\} = \lambda_0(t) \exp\{\alpha A + \beta L\}.$$

# Data-adaptive methods

- Consider the hazard function for the Cox PH model

$$\lambda\{t|A,L\} = \lambda_0(t)\exp\{\alpha A + \beta L\}.$$

- Data adaptive methods decide on reporting the $p$-value for the null hypothesis that $\alpha = 0$ in

$$\lambda\{t|A,L\} = \lambda_0(t)\exp\{\alpha A + \beta L\}$$

or the $p$-value for the null hypothesis that $\alpha_0 = 0$ in

$$\lambda\{t|A\} = \lambda_0^r(t)\exp\{\alpha_0 A\}$$

# Data-adaptive methods

- Consider the hazard function for the Cox PH model

$$\lambda\{t|A, L\} = \lambda_0(t) \exp\{\alpha A + \beta L\}.$$

- Data adaptive methods decide on reporting the $p$-value for the null hypothesis that $\alpha = 0$ in

$$\lambda\{t|A, L\} = \lambda_0(t) \exp\{\alpha A + \beta L\}$$

or the $p$-value for the null hypothesis that $\alpha_0 = 0$ in

$$\lambda\{t|A\} = \lambda_0^r(t) \exp\{\alpha_0 A\}$$

- One common strategy: adjust for $L$ iff significantly associated with outcome, conditional on exposure (e.g., at the 5% level)

# Censoring

- Not picking up certain variables is a consequence of large Type II errors

# Censoring

- Not picking up certain variables is a consequence of large Type II errors

- Suppose $L$ has a moderate effect on outcome, but a strong effect on censoring

# Censoring

- Not picking up certain variables is a consequence of large Type II errors
- Suppose $L$ has a moderate effect on outcome, but a strong effect on censoring
- Because censoring implies information loss and may even reduce variation in $L$ in the risk set, variable selection in the outcome model will rarely pick up $L$ when fitting

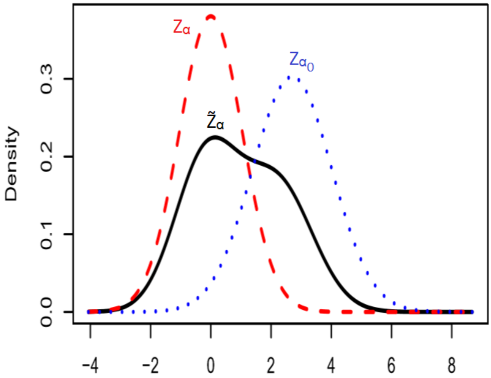
$$\lambda\{t|A, L\} = \lambda_0(t) \exp\{\alpha A + \beta L\}.$$

- Upon removing $L$ from the model, bias induced by informative censoring can result in highly inflated Type I errors

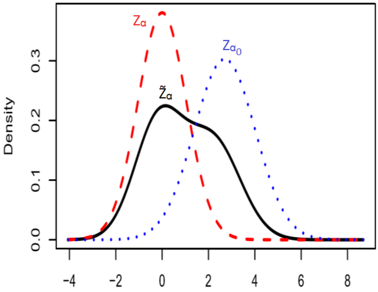
# Impact of Variable Selection

- Our "hesitation" whether or not to adjust for $L$ translates into a complex mixture distribution of the test statistic

$$\tilde{Z}_\alpha = \begin{cases} Z_\alpha & \text{if adjusted for } L \\ Z_{\alpha_0} & \text{if not adjusted for } L, \end{cases}$$



# Impact of Variable Selection

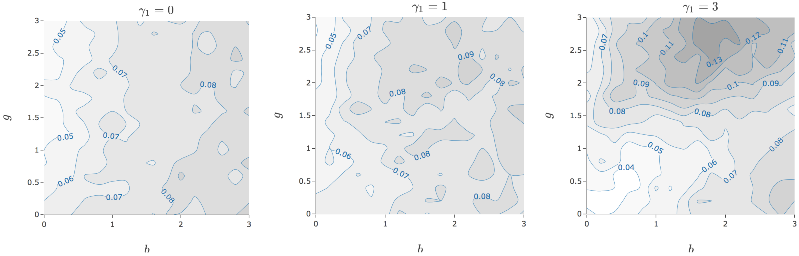



- Results from the test statistic jumping back and forth between $Z_\alpha$ and $Z_{\alpha_0}$
  - distribution of the latter might not be centered at zero

- This creates bias and inefficiency, and invalidates standard inference.

# Simulation: Test for Treatment Effect in RCT

Results obtained by post-Lasso




$n = 400$; $p = 30$; $A \stackrel{d}{=} Ber(0.5)$; $L \stackrel{d}{=} N(\mathbf{0}, \mathbb{I})$

$T \stackrel{d}{=} \exp(\lambda_T)$, with $\lambda_T = \exp(b \cdot \nu_T L)$ and $\nu_T = (1, 1/2, \ldots, 1/9, 1/10, 0_{11}, \ldots, 0_{30})'$

$C \stackrel{d}{=} \exp(\lambda_C)$, with $\lambda_C = \exp(\gamma_1 \cdot A + g \cdot \nu_C L)$ and
$\nu_C = (1, 1/2, \ldots, 1/5, 1, 1/2, \ldots, 1/5, 0_{11}, \ldots, 0_{30})'$

# How to account for censoring?

- For testing the null hypothesis of no treatment effect, we first perform a two stage selection procedure (e.g., by fitting two separate models using Lasso):

# How to account for censoring?

- For testing the null hypothesis of no treatment effect, we first perform a two stage selection procedure (e.g., by fitting two separate models using Lasso):

  1. The usual Cox model for survival time $T$, and

# How to account for censoring?

- For testing the null hypothesis of no treatment effect, we first perform a two stage selection procedure (e.g., by fitting two separate models using Lasso):

  1. The usual Cox model for survival time $T$, and
  2. a Cox model for censoring time $C$,

  both on exposure $A$ and baseline covariates $L$.

# How to account for censoring?

- For testing the null hypothesis of no treatment effect, we first perform a two stage selection procedure (e.g., by fitting two separate models using Lasso):

  1. The usual Cox model for survival time $T$, and
  2. a Cox model for censoring time $C$,

  both on exposure $A$ and baseline covariates $L$.

- Refit the Cox model for survival time, adjusting for covariates selected in either step:

$$\lambda\{t \mid A, L\} = \lambda_0^u(t)e^{\alpha_u A+\beta_u' L_u}.$$

# How to account for censoring?

- For testing the null hypothesis of no treatment effect, we first perform a two stage selection procedure (e.g., by fitting two separate models using Lasso):

  1. The usual Cox model for survival time $T$, and
  2. a Cox model for censoring time $C$,

  both on exposure $A$ and baseline covariates $L$.

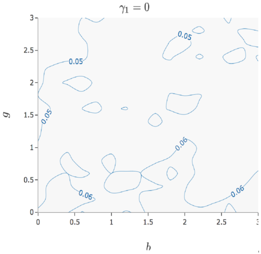
- Refit the Cox model for survival time, adjusting for covariates selected in either step:

$$\lambda\{t \mid A, L\} = \lambda_0^u(t)e^{\alpha_u A + \beta_u' L_u}.$$

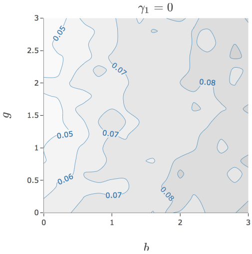
- Perform inference on $\alpha_u$ by conventional methods, based on robust SE (obtained via standard statistical software).

# Simulation: Test for Treatment Effect in RCT

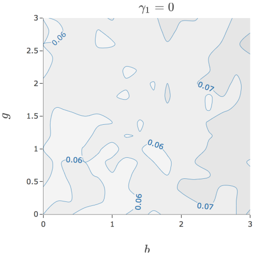
## Logrank




## Post-Lasso




## Poor Man's Method




$n = 400$; $p = 30$; $A \overset{d}{=} Ber(0.5)$; $L \overset{d}{=} N(\mathbf{0}, \mathbb{I})$

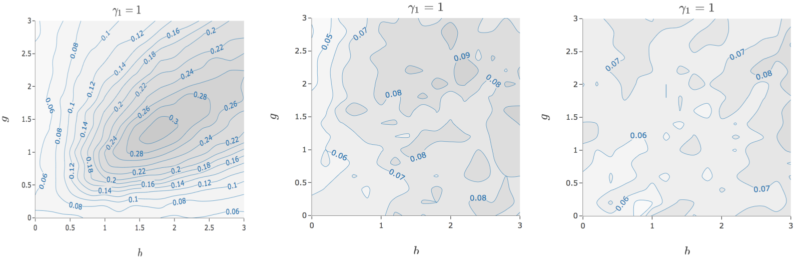
$T \overset{d}{=} \exp(\lambda_T)$, with $\lambda_T = \exp(b \cdot \nu_T L)$ and $\nu_T = (1, 1/2, \ldots, 1/9, 1/10, 0_{11}, \ldots, 0_{30})'$

$C \overset{d}{=} \exp(\lambda_C)$, with $\lambda_C = \exp(\gamma_1 \cdot A + g \cdot \nu_C L)$ and

$\nu_C = (1, 1/2, \ldots, 1/5, 1, 1/2, \ldots, 1/5, 0_{11}, \ldots, 0_{30})'$

# Simulation: Test for Treatment Effect in RCT

Logrank | Post-Lasso | Poor Man's Method




$n = 400$; $p = 30$; $A \overset{d}{=} Ber(0.5)$; $L \overset{d}{=} N(\mathbf{0}, \mathbb{I})$

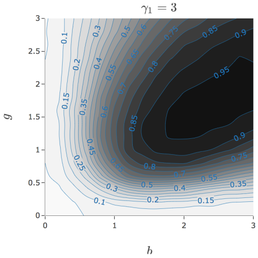
$T \overset{d}{=} \exp(\lambda_T)$, with $\lambda_T = \exp(b \cdot \nu_T L)$ and $\nu_T = (1, 1/2, \ldots, 1/9, 1/10, 0_{11}, \ldots, 0_{30})'$

$C \overset{d}{=} \exp(\lambda_C)$, with $\lambda_C = \exp(\gamma_1 \cdot A + g \cdot \nu_C L)$ and

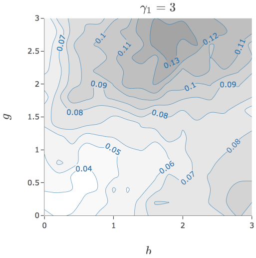
$\nu_C = (1, 1/2, \ldots, 1/5, 1, 1/2, \ldots, 1/5, 0_{11}, \ldots, 0_{30})'$

# Simulation: Test for Treatment Effect in RCT

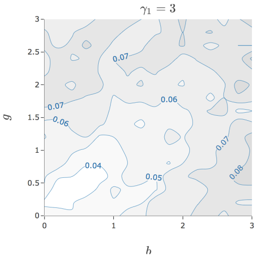
Logrank




Post-Lasso




Poor Man's Method




$n = 400$; $p = 30$; $A \stackrel{d}{=} Ber(0.5)$; $L \stackrel{d}{=} N(\mathbf{0}, \mathbb{I})$

$T \stackrel{d}{=} \exp(\lambda_T)$, with $\lambda_T = \exp(b \cdot \nu_T L)$ and $\nu_T = (1, 1/2, \ldots, 1/9, 1/10, 0_{11}, \ldots, 0_{30})'$

$C \stackrel{d}{=} \exp(\lambda_C)$, with $\lambda_C = \exp(\gamma_1 \cdot A + g \cdot \nu_C L)$ and
$\nu_C = (1, 1/2, \ldots, 1/5, 1, 1/2, \ldots, 1/5, 0_{11}, \ldots, 0_{30})'$

# My perspective on estimands, identification and estimation

**Estimand**

~~Model-free~~

**Identification**

Non-informative censoring **conditional on** $L$

**Estimation**

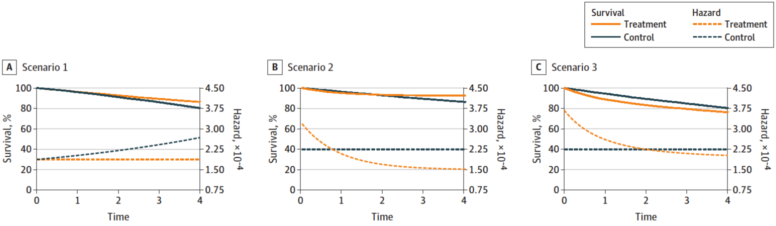
Double selection using Lasso

# Outline

# The proportional hazards assumption

- Hazards have been argued to be non-proportional in many settings.

Figure. Nonproportional Hazards and Survival Curves in 3 Hypothetical Trials Comparing a Treatment vs a Control

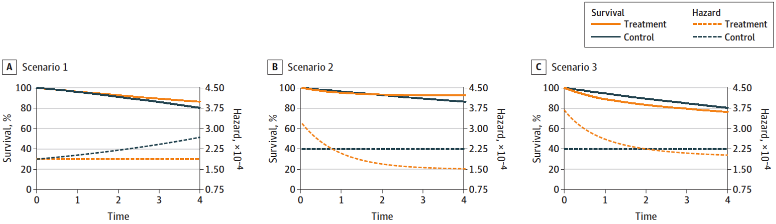



(Stensrud and Hernán, 2020)

# The proportional hazards assumption

- Hazards have been argued to be non-proportional in many settings.

Figure. Nonproportional Hazards and Survival Curves in 3 Hypothetical Trials Comparing a Treatment vs a Control




(Stensrud and Hernán, 2020)

What are we estimating in the Cox model when the proportional hazards assumption fails?

# Changing perspective

- We have defined our estimand as a parameter in a model.

# Changing perspective

- We have defined our estimand as a parameter in a model.
  - When the model is wrong:
    - No good understanding of what the partial likelihood estimator converges to.
    - What we infer depends on the estimator we use.
    - The target of the standard estimator depends on the censoring distribution. (Struthers and Kalbfleisch, 1986; Whitney et al., 2019)

# Changing perspective

- We have defined our estimand as a parameter in a model.
  - When the model is wrong:
    - No good understanding of what the partial likelihood estimator converges to.
    - What we infer depends on the estimator we use.
    - The target of the standard estimator depends on the censoring distribution. (Struthers and Kalbfleisch, 1986; Whitney et al., 2019)
  - This may not be of interest!

# Changing perspective

- We have defined our estimand as a parameter in a model.
  - When the model is wrong:
    - No good understanding of what the partial likelihood estimator converges to.
    - What we infer depends on the estimator we use.
    - The target of the standard estimator depends on the censoring distribution. (Struthers and Kalbfleisch, 1986; Whitney et al., 2019)
  - This may not be of interest!
- This highlights the benefits of choosing an estimand in a model-free way.
  - The estimand may coincide with the model parameter when assumptions hold...
  - ...but otherwise still captures the scientific question.

(van der Laan and Rose, 2011; Vansteelandt and Dukes, 2020)

# Application to the Cox model

- Reconsider the model

$$\lambda(t|A, L) = \lambda_0(t) \exp\{\alpha A + \beta L\}$$

- Ideally, we want an estimand that:
  - reduces to the log hazard ratio $\alpha$ when the model is correct.

# Application to the Cox model

- Reconsider the model

$$\lambda(t|A, L) = \lambda_0(t) \exp\{\alpha A + \beta L\}$$

- Ideally, we want an estimand that:
  - reduces to the log hazard ratio $\alpha$ when the model is correct.
  - is a weighted average of (log) causal hazard ratios when both parts of the model are wrong.

# Application to the Cox model

- Reconsider the model

$$\lambda(t|A, L) = \lambda_0(t) \exp\{\alpha A + \beta L\}$$

- Ideally, we want an estimand that:
  - reduces to the log hazard ratio $\alpha$ when the model is correct.
  - is a weighted average of (log) causal hazard ratios when both parts of the model are wrong.
  - does not depend on the censoring distribution.

# Application to the Cox model

- Reconsider the model

$$\lambda(t|A, L) = \lambda_0(t) \exp\{\alpha A + \beta L\}$$

- Ideally, we want an estimand that:
  - reduces to the log hazard ratio $\alpha$ when the model is correct.
  - is a weighted average of (log) causal hazard ratios when both parts of the model are wrong.
  - does not depend on the censoring distribution.

- Such estimands now exist.
  - Estimation methods allow for flexible machine learning methods.

(Whitney et al., 2019; Vansteelandt et al. 2022)

# My perspective on estimands, identification and estimation

| Estimand | Identification | Estimation |
|---|---|---|
| Model-free | Non-informative censoring **conditional on** $L$ | Machine learning |

# Outline

# Summary

- We often think of conditional causal effects as parameters in regression models.
- So long as we specify our estimand in advance, we have some freedom in letting the data choose our model, whilst maintaining type I error/interval coverage.

# Summary

- We often think of conditional causal effects as parameters in regression models.
- So long as we specify our estimand in advance, we have some freedom in letting the data choose our model, whilst maintaining type I error/interval coverage.
- Our estimand could be a regression parameter, or (even better) defined in a model-free way.
- The latter ensures that always return something that answers the question of interest.

Thank you for your attention!

E-mail: kelly.vanlancker@ugent.be
Website: kellyvanlancker.com

The opinions in this presentation are of the author and do not necessarily represent those of anyone else.

# Main references

Van Lancker, K., O. Dukes, S. Vansteelandt (2021). Principled selection of baseline covariates to account for censoring in randomized trials with a survival endpoint. Statistics in Medicine. DOI: 10.1002/sim.9017

Vansteelandt, S., O. Dukes, K. Van Lancker, T. Martinussen (2022). Assumption-Lean Cox Regression. Journal of the American Statistical Association. DOI: 10.1080/01621459.2022.2126362

Van Lancker, K., O. Dukes, S. Vansteelandt (2023). Ensuring valid inference for Cox hazard ratios after variable selection. Biometrics. DOI: 10.1111/biom.13889

# References I

Stensrud, M. J. and M. A. Hernán (2020).
Why test for proportional hazards?
*Jama 323*(14), 1401–1402.

Struthers, C. A. and J. D. Kalbfleisch (1986).
Misspecified proportional hazard models.
*Biometrika 73*(2), 363–369.

Van der Laan, M. J., S. Rose, et al. (2011).
*Targeted learning: causal inference for observational and experimental data*, Volume 4.
Springer.

Van Lancker, K., O. Dukes, and S. Vansteelandt (2021).
Principled selection of baseline covariates to account for censoring in randomized trials with a survival endpoint.
*Statistics in Medicine 40*(18), 4108–4121.

# References II


- Van Lancker, K., O. Dukes, and S. Vansteelandt (2023).
  Ensuring valid inference for cox hazard ratios after variable selection.

- Vansteelandt, S. and O. Dukes (2022).
  Assumption-lean inference for generalised linear model parameters.
  *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(3), 657–685.

- Vansteelandt, S., O. Dukes, K. Van Lancker, and T. Martinussen (2022).
  Assumption-lean cox regression.
  *Journal of the American Statistical Association*, 1–10.

# References III

Whitney, D., A. Shojaie, and M. Carone (2019).
Comment: Models as (deliberate) approximations.
*Statistical science: a review journal of the Institute of Mathematical Statistics 34*(4), 591.