FACULTY OF SCIENCES

A GENTLE INTRODUCTION TO TARGETED LEARNING IN RCTS: WHAT, WHY AND HOW?

DATA-ADAPTIVE INFERENCE FOR CONDITIONAL CAUSAL EFFECTS

Oliver Dukes, Ghent University

GHENT UNIVERSITY

Acknowledgements

Thanks to

- Kelly Van Lancker (UGent) for organising this webinar.
- Kelly and Stijn Vansteelandt (UGent, LSHTM) for their collaboration.
- The Ghent University Special Research Fund (BOF).

INTRODUCTION

GHENT
UNIVERSITY

Conditional effect estimands

- Conditional effects describe the impact of a treatment within subpopulations who share the same covariate values.

CONDITIONAL EFFECT ESTIMANDS

- Conditional effects describe the impact of a treatment within subpopulations who share the same covariate values.
- This can mean e.g. separately estimating

$$E(Y^1 - Y^0|\text{Sex}=\text{F}) \quad \text{and} \quad E(Y^1 - Y^0|\text{Sex}=\text{M})$$

What is the treatment effect in women? And in men?

Conditional effect estimands

- Conditional effects describe the impact of a treatment within subpopulations who share the same covariate values.
- This can mean e.g. separately estimating

$$E(Y^1 - Y^0|\text{Sex=F}) \quad \text{and} \quad E(Y^1 - Y^0|\text{Sex=M})$$

  What is the treatment effect in women? And in men?
- Due to non-collapsibility, a conditional estimand assumed to be homogeneous may still differ from the corresponding marginal estimand.
- Even if

$$\phi = \frac{P(Y^1 = 1|W)/P(Y^1 = 0|W)}{P(Y^0 = 1|W)/P(Y^0 = 0|W)}$$

  does not vary with W, it $\neq$ the marginal odds ratio (in general).

TARGETED LEARNING OF CONDITIONAL EFFECTS

- Compared with marginal estimands,
  targeted learning for conditional effects is more subtle.
- Conventional thinking ties these effects to parametric models.

TARGETED LEARNING OF CONDITIONAL EFFECTS

- Compared with marginal estimands,
  targeted learning for conditional effects is more subtle.
- Conventional thinking ties these effects to parametric models.
- Without some simplifying assumptions
  (e.g. effect homogeneity), conditional effects like

$$\frac{P(Y^1 = 1|W)/P(Y^1 = 0|W)}{P(Y^0 = 1|W)/P(Y^0 = 0|W)}$$

  may be high-dimensional and difficult to report.
- Statistical inference is also very challenging.

Outline of this talk

- I will first explore how to obtain valid inference for conditional causal effects, in combination with variable selection.

Outline of this talk

- I will first explore how to obtain valid inference for conditional causal effects, in combination with variable selection.
- I will focus on the hazard ratio.

Outline of this talk

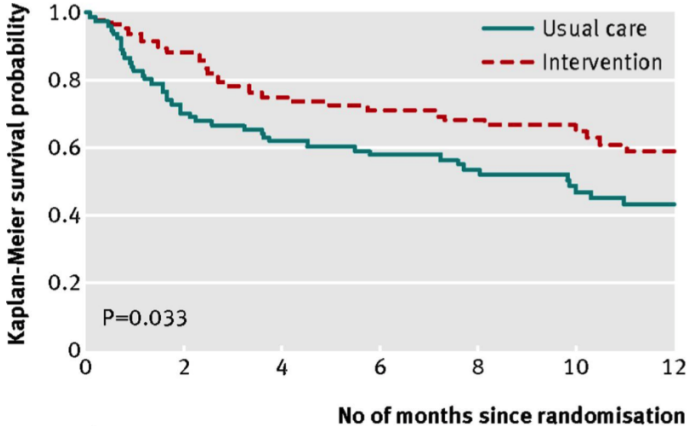
- I will first explore how to obtain valid inference for conditional causal effects, in combination with variable selection.
- I will focus on the hazard ratio.
- The proposal relies on potentially restrictive assumptions (e.g. proportional hazards).
- Closer to the spirit of targeted learning, I will discuss ongoing work on inferring conditional causal hazard ratios whilst relaxing model assumptions as far as possible.

INFERENCE FOR THE HAZARD RATIO

GHENT
UNIVERSITY

TIME-TO-EVENT OUTCOMES

- When the outcome is a time-to-event, we have many options for describing the data.



(Sedgwick and Joekes, 2013)

- Survival curves offer a convenient visual aid.

SUMMARISING THE TREATMENT EFFECT

- Suppose we want to use these curves to obtain a simple summary of the treatment effect (+ confidence intervals).
- I could compare the chance of surviving to month 2 between treatment and control groups.
  - Or surviving to month 4....
  - ...or month 6....or 8....

SUMMARISING THE TREATMENT EFFECT

- Suppose we want to use these curves to obtain a simple summary of the treatment effect (+ confidence intervals).
- I could compare the chance of surviving to month 2 between treatment and control groups.
  - Or surviving to month 4....
  - ...or month 6....or 8....
- When measuring associations w.r.t event time,
  the strength of the association can depend upon the chosen window of time.
- This also applies to other common measures (e.g. restricted mean survival time).

The hazard ratio

- Unlike competing measures,
  a constant hazard ratio does not require pre-specification of a time window.
- In addition to other nice properties of conditional estimands,
  a (conditional) hazard ratio thus offers
  a convenient summary of the association of interest.

THE HAZARD RATIO

- Unlike competing measures,
  a constant hazard ratio does not require pre-specification of a time window.
- In addition to other nice properties of conditional estimands,
  a (conditional) hazard ratio thus offers
  a convenient summary of the association of interest.
- Things became more complicated when hazards are non-proportional....

THE COX MODEL ADJUSTED FOR COVARIATES

- For the moment, I will assume that there is no censoring.
- We might fit the adjusted Cox model:

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- This model expresses:
  - the treatment-hazard association.
  - the covariate-hazard association.

The Cox model adjusted for covariates

- For the moment, I will assume that there is no censoring.
- We might fit the adjusted Cox model:

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- This model expresses:
  - the treatment-hazard association.
  - the covariate-hazard association.
- If both the red and blue parts are correct,
  then the partial likelihood estimator of β_1
  captures the conditional causal (log) hazard ratio.

The Cox model adjusted for covariates

- For the moment, I will assume that there is no censoring.
- We might fit the adjusted Cox model:

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- This model expresses:
  - the treatment-hazard association.
  - the covariate-hazard association.
- If both the red and blue parts are correct,
  then the partial likelihood estimator of β_1
  captures the conditional causal (log) hazard ratio.
- Data-driven variable selection tempers concerns about misspecification.

How can we use the data to select a model? (1)

Let's look first at the part involving the treatment effect.

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- Assumption: the conditional hazard ratio for treatment doesn't depend on time, or on covariates.
- One might test for interactions, or use the Lasso to check this.

How can we use the data to select a model? (1)

Let's look first at the part involving the treatment effect.

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- Assumption: the conditional hazard ratio for treatment doesn't depend on time, or on covariates.
- One might test for interactions, or use the Lasso to check this.
- However, standard post-selection p-values/confidence intervals for β_1 are invalid....

How can we use the data to select a model? (2)

Suppose we used the data to select the blue part of the model,
not involving treatment.

- Cox model + stepwise selection strategy to select covariates.
- How does this effect inference on the target of interest e.g. β_1?

How can we use the data to select a model? (2)

Suppose we used the data to select the blue part of the model,
not involving treatment.

- Cox model + stepwise selection strategy to select covariates.
- How does this effect inference on the target of interest e.g. β_1?
- Tests of the null hypothesis that β_1=0 remain valid,
  if we ask for a robust standard error.

  (Van Lancker et al., 2020)
- This is also extends to tests of the null in generalised linear models.

  (Belloni et al., 2014, 2016)

How can we use the data to select a model? (2)

Suppose we used the data to select the blue part of the model,
not involving treatment.

- Cox model + stepwise selection strategy to select covariates.
- How does this effect inference on the target of interest e.g. β_1?
- Tests of the null hypothesis that β_1=0 remain valid,
  if we ask for a robust standard error.

  (Van Lancker et al., 2020)
- This is also extends to tests of the null in generalised linear models.

  (Belloni et al., 2014, 2016)
- It is also *approximately* the case for confidence intervals for β_1
  when the treatment effect is small in magnitude.
- Valid intervals can be obtained with an additional weighted regression step.

How can we use the data to select a model? (2)

Suppose we used the data to select the blue part of the model,
not involving treatment.

- Cox model + stepwise selection strategy to select covariates.
- How does this effect inference on the target of interest e.g. β_1?
- Tests of the null hypothesis that β_1=0 remain valid,
  if we ask for a robust standard error.

  (Van Lancker et al., 2020)
- This is also extends to tests of the null in generalised linear models.

  (Belloni et al., 2014, 2016)
- It is also *approximately* the case for confidence intervals for β_1
  when the treatment effect is small in magnitude.
- Valid intervals can be obtained with an additional weighted regression step.

Inference after variable selection

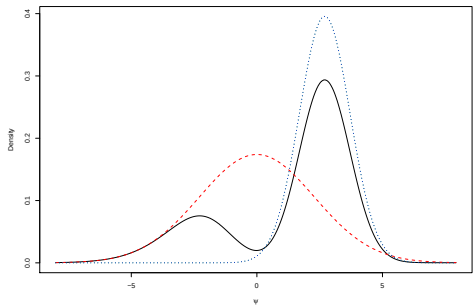
Data-adaptive inference is easier when our target (e.g. β_1) is chosen in advance.

WHY DOES THIS WORK?

- In observational data, W may include confounders.
- In selecting confounders based on their association with an outcome, we may miss variables weakly predictive of the outcome, but strongly associated with A.

WHY DOES THIS WORK?

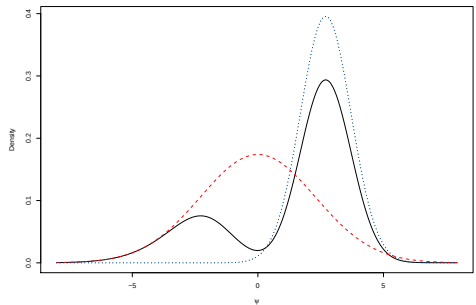
- In observational data, W may include confounders.
- In selecting confounders based on their association with an outcome, we may miss variables weakly predictive of the outcome, but strongly associated with A.



(Dukes and Vansteelandt, 2020)

WHY DOES THIS WORK?

- In observational data, W may include confounders.
- In selecting confounders based on their association with an outcome, we may miss variables weakly predictive of the outcome, but strongly associated with A.



(Dukes and Vansteelandt, 2020)

- If treatment is randomised, then these concerns disappear....

CENSORING COMPLICATES MATTERS

- With informative censoring, we may have no option other than adjusting for covariates.

CENSORING COMPLICATES MATTERS

- With informative censoring, we may have no option other than adjusting for covariates.
- Suppose we test for whether to adjust for W_1 in our Cox model.

CENSORING COMPLICATES MATTERS

- With informative censoring, we may have no option other than adjusting for covariates.
- Suppose we test for whether to adjust for W_1 in our Cox model.
- If we keep it, we implicitly assume that censoring is independent of survival time given treatment and W_1.
- If we remove it, we make a stronger censoring assumption.

CENSORING COMPLICATES MATTERS

- With informative censoring, we may have no option other than adjusting for covariates.
- Suppose we test for whether to adjust for W_1 in our Cox model.
- If we keep it, we implicitly assume that censoring is independent of survival time given treatment and W_1.
- If we remove it, we make a stronger censoring assumption.
- Ignoring this can lead to tests of the causal null having inflated Type I error.

Censoring and variable selection

Changing the adjustment set changes our censoring assumption!

How to account for censoring?

- For testing the null hypothesis of no treatment effect, Van Lancker et al. (2020) propose the following:
  1. Perform selection based on a Cox model for survival time.

How to account for censoring?

- For testing the null hypothesis of no treatment effect, Van Lancker et al. (2020) propose the following:
  1. Perform selection based on a Cox model for survival time.
  2. Perform a second selection step based on a Cox model for censoring time.

How to account for censoring?

- For testing the null hypothesis of no treatment effect, Van Lancker et al. (2020) propose the following:
  1. Perform selection based on a Cox model for survival time.
  2. Perform a second selection step based on a Cox model for censoring time.
  3. Refit the Cox model for survival time, adjusting for covariates selected in either step.

How to account for censoring?

- For testing the null hypothesis of no treatment effect,
  Van Lancker et al. (2020) propose the following:
  1. Perform selection based on a Cox model for survival time.
  2. Perform a second selection step based on a Cox model for censoring time.
  3. Refit the Cox model for survival time, adjusting for covariates selected in either step.
  4. Test $\beta_1=0$ based on the final model.

How to account for censoring?

- For testing the null hypothesis of no treatment effect, Van Lancker et al. (2020) propose the following:
  1. Perform selection based on a Cox model for survival time.
  2. Perform a second selection step based on a Cox model for censoring time.
  3. Refit the Cox model for survival time, adjusting for covariates selected in either step.
  4. Test $\beta_1=0$ based on the final model.
- This procedure is seen to preserve Type I error when survival times are independent of censoring times, given A and W.
- Variables that are likely only predictive of censoring should be removed in advance of seeing the data.

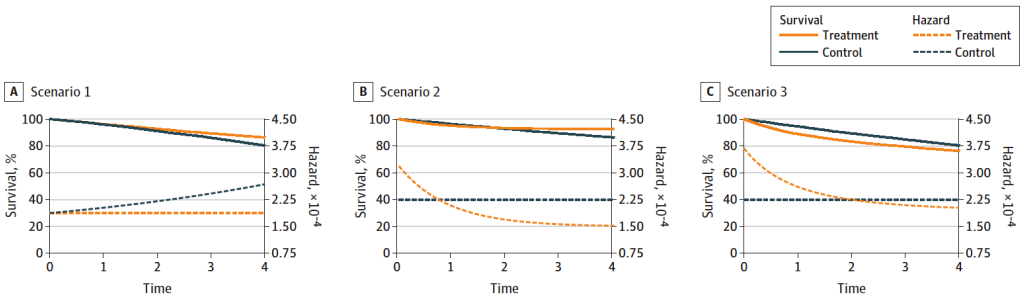
See Alex's talk

Relaxing the assumptions

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'S THE TARGET?

- What are we estimating in the Cox model when the proportional hazards assumption fails?
- No good understanding of what the partial likelihood estimator converges to when the model is wrong.....
- Under misspecification, the target of the standard estimator depends on the censoring distribution.

  (Struthers and Kalbfleisch, 1986; Whitney et al., 2019)

- The situation becomes even more complicated with covariates.

CHANGING PERSPECTIVE

- We have defined our estimand as a parameter in a model.
  - When the model is wrong, what we infer depends on the estimator we use.
  - This may not be of interest!

CHANGING PERSPECTIVE

- We have defined our estimand as a parameter in a model.
  - When the model is wrong, what we infer depends on the estimator we use.
  - 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

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- Ideally, we want an estimand that:
  - reduces to the log hazard ratio β_1 when the model is correct.

Application to the Cox model

- Reconsider the model

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- Ideally, we want an estimand that:
  - reduces to the log hazard ratio β_1 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

$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$

- Ideally, we want an estimand that:
  - reduces to the log hazard ratio β_1 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
$$\lambda(t|A, W) = \lambda_0(t) \exp\{\beta_1 A + \beta_2^T W\}$$
- Ideally, we want an estimand that:
  - reduces to the log hazard ratio β_1 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.

  (Whitney et al., 2019; Vansteelandt et al. 2021)
- Whitney et al. (2019) weight by the marginal time-to-event distribution; more efficient choices are also available.

Estimation

- Inferring this new estimand
  requires estimation of the conditional hazard (given A and W),
  as well as the censoring mechanism.
- This can be done using variable selection,
  but also more flexible machine learning methods.

Estimation

- Inferring this new estimand
  requires estimation of the conditional hazard (given A and W),
  as well as the censoring mechanism.
- This can be done using variable selection,
  but also more flexible machine learning methods.
- As for the approaches in Alex's talk,
  we can still obtain valid tests and confidence intervals afterwards.

Estimation

- Inferring this new estimand
  requires estimation of the conditional hazard (given A and W),
  as well as the censoring mechanism.
- This can be done using variable selection,
  but also more flexible machine learning methods.
- As for the approaches in Alex's talk,
  we can still obtain valid tests and confidence intervals afterwards.
- Even if the selected model is highly complex,
  we still return a scalar summary of the association of interest.

SUMMARY

GHENT
UNIVERSITY

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.

SOFTWARE

- See the ‘hdm’ package in R, for valid confidence intervals and hypothesis tests for conditional treatment effects indexing linear and logistic models, after variable selection using the Lasso.

  (Chernozhukov et al., 2016)

- Much more to be done for time-to-event settings.

References I

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Belloni, A., Chernozhukov, V., and Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34(4):606–619.

Chernozhukov, V., Hansen, C., and Spindler, M. (2016). hdm: High-dimensional metrics. *arXiv preprint arXiv:1608.00354*.

Dukes, O. and Vansteelandt, S. (2020). How to obtain valid tests and confidence intervals after propensity score variable selection? *Statistical methods in medical research*, 29(3):677–694.

Sedgwick, P. and Joekes, K. (2013). Kaplan-meier survival curves: interpretation and communication of risk. *Bmj*, 347.

References II

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

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

van der Laan, M. J. and Rose, S. (2011). *Targeted Learning*. Springer Series in Statistics. Springer New York, New York, NY.

Van Lancker, K., Dukes, O., and Vansteelandt, S. (2020). Principled selection of baseline covariates to account for censoring in randomized trials with a survival endpoint. *arXiv preprint arXiv:2007.08190*.

Vansteelandt, S. and Dukes, O. (2020). Assumption-lean inference for generalised linear model parameters. *arXiv preprint arXiv:2006.08402*.

References III

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