

Causal inference and survival analysis

Torben Martinussen

Department of Biostatistics

October 5, 2021

Todays plan

- Basic causal inference.
- Causal interpretation of the hazard ratio.
- Other measures to report that has a causal interpretation.
- Basic instrumental variable (IV) techniques.
- IV and additive hazards I.
- IV and additive hazards II.
- IV and Cox regression.

What do we mean by "causal inference"?

- Much of the analysis of data in health and social sciences has as its central aim the quest to learn about cause-effect relationships.
- Does this treatment work? How harmful is the exposure?
- These are causal questions.
- Randomised studies can answer such questions.
- But sometimes we cannot do the randomization, or we have data from an observational study.
- The goal of many observational studies is still to make inferences about the effects of causes.



Patients with liver metastases from colorectal cancer

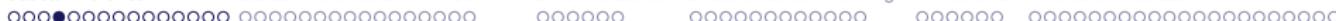
cancer (Kemeny et al., 2002). Patients, accrued between August 1990 and January 1997, were randomly assigned to surgical resection alone (standard arm S) or surgical resection followed by chemotherapy (intervention arm I). Of the 109 patients, 56 (53) were randomized to arm S (arm I). Our aim is to compare the five-year survival chance when treated with resection and chemotherapy versus resection alone.

The chemotherapy was administered through an arterial device (abbreviated AD), implanted at the time of surgical resection. For reasons likely related to survival, the operating surgeons sometimes decided not to implant an AD in patients on arm I. Ten such nonreceivers fell back onto the standard care of arm S, while 43 did receive an AD. Figure 1(b) shows Kaplan-Meier curves for both groups. The observed difference does not necessarily indicate a beneficial treatment. Indeed, if noncompliance were selective and people with worse prognosis complied less, this could follow even from a harmful treatment effect. Similarly, the as-treated comparison (Figure 1(c)) is



Example

- The level of the hormone AMH reflects the fertility of a given woman.
 - We would like to investigate the effect of taking contraceptive pill.
 - We have observations for 732 women.
 - Y is level of AMH
 - A is taking the pill (yes/no).
 - AGE, smoking, bmi?



Example

- How would we analyse these without the aid of causal inference thinking?
 - We might start by looking at the difference in means (using log-scale for amh):

```

> mean(log_amh[ppiller==0])
[1] 2.911799
> mean(log_amh[ppiller==1])
[1] 2.853779
> summary(lm(log_amh~factor(ppiller)))
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.91180   0.03739  77.872 <2e-16 ***
factor(ppiller)1 -0.05802   0.06700  -0.866   0.387

```

What is the estimand here?

$$E(Y|A=1) - E(Y|A=0)$$

Example

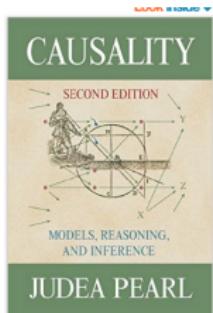
- Intuitively we know we can't give this a causal interpretation as there is likely to be confounding.
- Age is known to affect the level of amh, and the age distribution may be different for the pill-users compared to non-users.
- Sometimes this is called confounding bias
- But if so, bias relative to what?
- We must then have another estimand in mind, other than $E(Y|A = 1) - E(Y|A = 0)$
- So what is it?
- Without a mathematical vocabulary even to write down the estimand we are interested in, there is very little hope of rigorous progress ...

Causal inference

- Probability and statistics lets us describe aspects of the joint distribution of the variables observed in our dataset: means, variances, regression parameters, ...
- But this standard language does not include a vocabulary for expressing how this distribution would change in response to an external intervention.
- What would the mean of amh be if, contrary to reality, every women were forced to take the pill.
- What would the mean of amh be if, contrary to reality, every women were forbidden to take the pill.

Causal inference

- There are two ways of expressing such notions.
- One is due to Judea Pearl, do-calculus
- The other approach (which we follow) uses so-called counterfactuals.
- Was introduced by Neyman in 1923 and developed by Rubin in 1974.
- But much of the modern development of causal inference using counterfactuals is due to Jamie Robins.



Counterfactuals

- So what is a counterfactual (or potential outcome).
- The idea is: Each subject in the target population has two potential outcomes:
- $Y^{a=0}$ the outcome if allocated to treatment $a = 0$
- $Y^{a=1}$ the outcome if allocated to treatment $a = 1$
- At least one is counterfactual and will never be realized, since subject receives at most one of the treatments.

Example: Greek gods

- 20 greek gods heart patients.
- Treatment: Heart transplant (yes/no). Denote it A (values 1/0).
- Outcome: five day survival (alive/dead). Denote it Y (values 1/0).
- $Y^{a=1}$ denote what the outcome would have been under treatment value $a = 1$
- $Y^{a=0}$ denote what the outcome would have been under treatment value $a = 0$
- Only one of these is observed! We call Y^1 and Y^0 for **potential outcomes**, or **counterfactual outcomes**.
- Consistency assumption: If $A = a$ then $Y^a = Y$. THIS LINKS COUNTERFACTUALS TO THE OBSERVED DATA.

Example: Greek gods

Table 1.1

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Cyclope	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

- Average causal effect if: $E(Y^1) \neq E(Y^0)$
- For binary outcome Y , $E(Y) = P(Y = 1)$.
- Is there an average causal effect here?
- Problem is that we cannot observe Table 1.1!

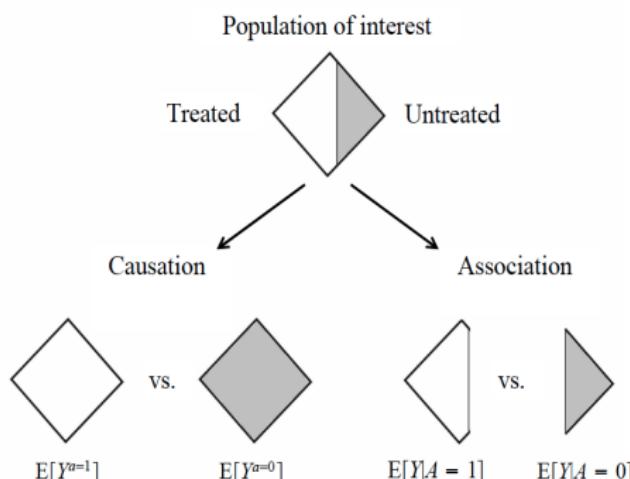
Causation versus association

Table 1.2

	<i>A</i>	<i>Y</i>
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Cyclope	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

- We never get to see all counterfactuals.
- Table 1.2 gives the observed data.
- Calculate from Table 1.2 $P(Y = 1|A = 1)$ and $P(Y = 1|A = 0)$.
- There seems to be an association betw. Y and A .
- Association is not causation!

Causation versus association



- Association is defined by a different risk in two disjoint subsets of the population defined subjects actual treatment.
- Causation is a contrast betw. the whole population all being treated and the same population all being untreated.

Exchangeability

- How can we ever estimate a causal effect when we never get to see all counterfactuals?
- We can estimate the causal effect under so-called exchangeability:

$$Y^a \coprod A$$

for all a .

- Also called the no unmeasured confounders assumption.
- In words, the treated and the untreated would have experienced the same risk of death if they had received the same treatment (either $a = 0$ or $a = 1$): they are exchangeable.
- In a randomized study $P(A = 1)$ does not depend on anything (treatment is randomized!), so exchangeability holds!
- Y^a is latent.



Exchangeability

- Why does exchangeability mean that we can estimate a causal effect?
- This holds, because

$$E(Y^1) - E(Y^0) = E(Y^1|A=1) - E(Y^0|A=0) = E(Y|A=1) - E(Y|A=0)$$

so here association is causation!

- This is why randomized studies are so important!
- Generally we are unable to check whether exchangeability holds in practice (unless the study was randomized).

Conditionally exchangeability

- Conditionally exchangeability $Y^a \perp\!\!\!\perp A|W$.
- What we just did can also be written as

$$P(Y^a = 1) = \sum_I P(Y = 1|A = a, W = w)P(W = w) \quad (1)$$

and is also called standardization.

- (1) is also called G-formula.
- We can also write it for the mean as:

$$E(Y^a) = \frac{1}{n} \sum_i E(Y|A = a, W_i) \quad (2)$$

Causal Inference

Causal Inference and the Cox HR

Other measures

Unobserved confounding

IV-methods

Time-to-event data



Subtleties in the interpretation of the Cox HR

Survival analysis

- We will study time-to-event data (death, onset of disease).
- We let T denote this timing.
- We allow for right-censoring, letting C denote the censoring time.
- We consider explanatory variables (A, W), A indicates treatment. Leave out the W to begin with.
- Traditionally such data has often been modelled using the hazard function (for instance Cox model)

$$\lambda_T(t|A, V) = \lambda_0(t)e^{\beta_A A}$$

- Let's now assume we have a randomized study with treatment indicator A .
- A careful check of data suggest that

$$\lambda(t; A) = \{e^{\beta_1 A} I(t \leq 4) + e^{\beta_2 A} I(t > 4)\} \lambda_0(t)$$

fits perfectly. One HR on $[0, 4]$ and another one on $(4, \tau[$.

Survival analysis

Remember,

$$\lambda(t; A) = \{e^{\beta_1 A} I(t \leq 4) + e^{\beta_2 A} I(t > 4)\} \lambda_0(t)$$

fits data perfectly.

Analysis performed in R:

Call:

```
coxph(formula = Surv(tstart, time, status) ~ A * strata(timegroup),
      data = data2)
```

	coef	exp(coef)	se(coef)	z	p
A	-0.6692	0.5121	0.0183	-36.6	<2e-16
A:strata(timegroup)timegroup=2	0.6385	1.8936	0.0395	16.2	<2e-16

[1] "Estimate in second interval and s.e."

[1] -0.03073876 0.03495531

- Estimates: $e^{\hat{\beta}_1} = 0.51$ and $e^{\hat{\beta}_2} = e^{-0.03} = 0.97$ (0.91,1.04).
- Conclusion about treatment effect?

Survival analysis and causal inference

- Are the above calculations useful in any respect?
- In a randomized study, can we say something about time-changing treatment effect?
- Using HR's?
- Using the model (say the doctor/substance matter informed us about the change point v , eg =4)

$$\frac{\lambda(t|A=1)}{\lambda(t|A=0)} = \begin{cases} e^{\beta_1} & \text{if } t \leq v \\ e^{\beta_2} & \text{if } t > v \end{cases}$$

Survival analysis and causal inference

- NO



- A HR cannot be given a causal interpretation (unless it's 1); even in a randomized study!
- Hazards of hazard ratios (Hernan, Epidemiology, 2010).

Survival analysis and causal inference

- Causal effect of randomized binary exposure variable A on a survival time response T in the presence of a unmeasured variable Z (selection) that also has an effect on T .
- A and Z are independent due to randomization.
- Let $P(T^a > t)$ have hazard function $\lambda(t; A)$. Will explore if the true DGP, $\lambda(t; A, Z)$ can be chosen such that

$$\frac{\lambda(t; A = 1)}{\lambda(t; A = 0)} = \begin{cases} e^{\beta_1} & \text{if } t \leq v \\ e^{\beta_2} & \text{if } t > v \end{cases}$$

with $\beta_1 \neq \beta_2$.

- We can estimate these parameters un-biasedly due to the randomization.
- Write $\lambda(t; A) = \{e^{\beta_1 A} I(t \leq v) + e^{\beta_2 A} I(t > v)\} \lambda_0(t)$

Exercise

Assume that the survival time T has the conditional hazard function

$$\lambda(t|A, Z) = Z\lambda_0(t)e^{\beta A} \exp\{\theta\Lambda_0(t)e^{\beta A}\}$$

given A and Z , where A is binary (treatment indicator, $A = 1$ corresponds to active treatment). The so-called frailty variable Z is unobserved and assumed to be Gamma distributed with mean 1 and variance θ . It also assumed that $\lambda_0(t)$ is hazard function and that $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$. The term "frailty" stems from the fact that the larger Z the larger $\lambda(t|A, Z)$ (ie more frail). We also assume that A and Z are independent, which will be the case if treatment is randomized.

- (a) Derive the observed hazard function, that is, $\lambda(t|A)$, and show that

$$\frac{\lambda(t|A=1)}{\lambda(t|A=0)} = \exp(\beta) \quad (3)$$

The model in (1) is called a Cox-model (proportional hazards). Assume in the following that $\beta < 0$ (beneficial treatment effect).

- (b) Show that

$$\frac{E(Z|T > t, A=1)}{E(Z|T > t, A=0)} = \exp\{\Lambda_0(t)(1 - e^\beta)\}$$

- (c) Use the result in (b) to conclude about the individuals in the two treatment groups concerning whether they are equally "frail" as time passes by. Start with considering $t = 0$.

Survival analysis and causal inference

- We will assume that $\lambda(t; a, z) = z\lambda^*(t; a)$, and let $Z \sim \Gamma(1/\theta, \theta)$ with $\theta = 1$ to make the calculations simple.
- Let $\phi_Z(u) = E(e^{-Zu})$ be the Laplace transform associated with the distribution of Z .
- One can derive the following relationship between the hazard function of interest $\lambda(t; A)$, and $\lambda^*(t; A)$ (formulated via the cumulated quantities),

$$\Lambda^*(t|a) = \phi_Z^{-1}(e^{-\Lambda(t;a)}) = \frac{1 - e^{-\Lambda(t;a)}}{e^{-\Lambda(t;a)}}$$

- Since we have decided on the structure on $\Lambda(t; a)$ (Cox with change point), we can find $\Lambda^*(t; a)$ by a direct calculation using the expression in the latter display.
- Simple calculations then leads to

$$\lambda(t; A, Z) = \begin{cases} Z\lambda_0(t)e^{\beta_1 A} \exp\{\Lambda_0(t)e^{\beta_1 A}\} & \text{if } t \leq v \\ Z\lambda_0(t)e^{\beta_2 A} \exp\{\Lambda_0(v)e^{\beta_1 A} + \Lambda_0(v, t)e^{\beta_2 A}\} & \text{if } t > v \end{cases}$$

where $\Lambda_0(v, t) = \int_v^t \lambda_0(s) ds$.

Survival analysis and causal inference

- Now, let

$$\text{HR}_Z(t) = \frac{\lambda(t; A = 1, Z)}{\lambda(t; A = 0, Z)}$$

be the true (?) hazard ratio.

- Then

$$\text{HR}_Z(t) = \begin{cases} e^{\beta_1} \exp [\Lambda_0(t)\{e^{\beta_1} - 1\}] & \text{if } t \leq v \\ e^{\beta_2} \exp [\Lambda_0(v)\{e^{\beta_1} - 1\} + \Lambda_0(v, t)\{e^{\beta_2} - 1\}] & \text{if } t > v \end{cases}$$

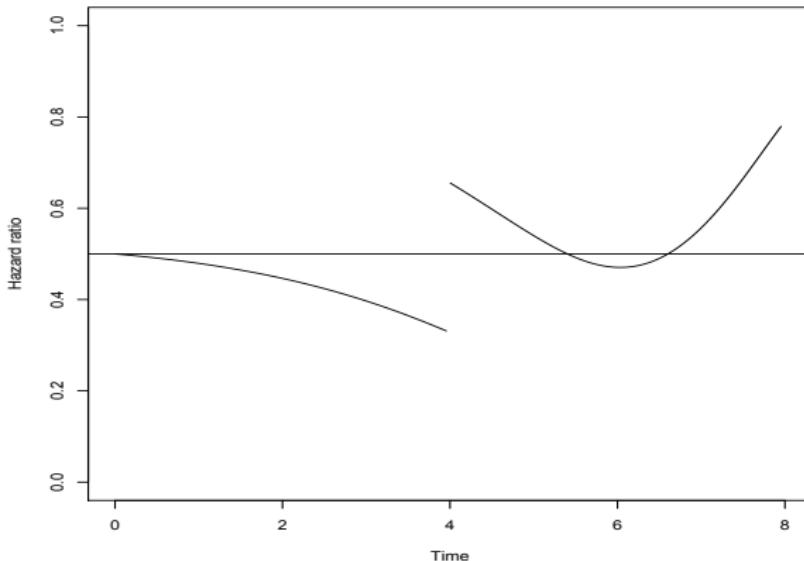
- Assume that $\beta_1 < 0$, and $\beta_2 = 0$.
- The above expression for $\text{HR}_Z(t)$ then simplifies to

$$\text{HR}_Z(t) = \begin{cases} e^{\beta_1} \exp [\Lambda_0(t)\{e^{\beta_1} - 1\}] & \text{if } t \leq v \\ \exp [\Lambda_0(v)\{e^{\beta_1} - 1\}] & \text{if } t > v \end{cases}$$

- Since $e^{\beta_1} < 1$ we see that $\text{HR}_Z(t) < 1$, that is, there is a beneficial effect of the treatment at all times!
- However, the conclusion based on the change point Cox model would be that the treatment effect disappears after time v since $e^{\beta_2} = 1$

Survival analysis and causal inference

$\text{HR}_Z(t)$ when $\beta_1 = \log(1/2)$, $\beta_2 = 0$ and $\nu = 4$:



SO BENEFICIAL TREATMENT EFFECT THROUGHOUT!

Survival analysis and causal inference

- Keep in mind that we do not obtain biased estimates, but arguing about the treatment effect using the hazard ratio **simply leads to a wrong conclusion**.
- We have, for $t \leq v$, that

$$E(Z|T > t, A = a) = \begin{cases} \exp\{-\Lambda_0(t)\} & \text{if } a = 0 \\ \exp\{-\Lambda_0(t)e^{\beta_1}\} & \text{if } a = 1 \end{cases}$$

so indeed selection is taking place:

$$E(Z|T > t, A = 1) > E(Z|T > t, A = 0).$$

- Comparing $P(T^a > t)$, $a = 0, 1$ will not give wrong answers but these are not very useful to conclude anything about time-dynamics.

Survival analysis and causal inference

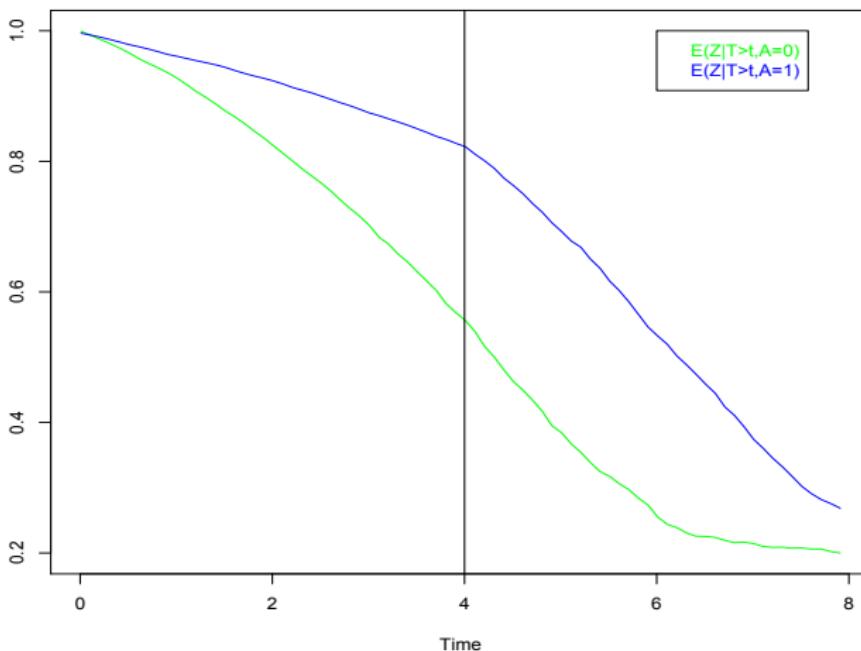


Figure: Plot of $E(Z|T > t, A = a)$, $a = 0, 1$, $n = 20000$.

Conclusions so far

- Never report a HR if you want to express a causal effect of treatment! At least if you interpret the HR as a HR.



- But there are so many papers doing exactly that !!!???
- Jonathan Bartlett just wrote me "In 2018 a workshop was held in the US (Duke) which discussed outputs from a working group composed of statisticians from many pharmaceutical companies on the topic of non-proportional hazards in clinical trials. You can watch (8 hours!) the workshop on Youtube: <https://youtu.be/npufYAHcoxk> I discovered today that their work was published a couple of months ago in "Statistics in Biopharmaceutical Research", a copy of which I attach."
- Show p. 9-10 in paper referred to by Bartlett.
- Can we ever estimate dynamic effects? Here $P(T \leq t)$ isn't very helpful either.

Cox-regression and causal inference

- Cox hazard ratio does not have a causal interpretation:

$$\phi = \frac{P(T^1 = t | T^1 \geq t)}{P(T^0 = t | T^0 \geq t)}$$

This is where you can see the key problem.

- A hazard ratio that does have a causal interpretation is

$$\text{HR}(t) = \frac{P(T^1 = t | T^0 \geq t, T^1 \geq t)}{P(T^0 = t | T^0 \geq t, T^1 \geq t)}$$

We can never estimate $\text{HR}(t)$ from the observed data.

- But can we show, under some positive correlation (?) btw T^0 and T^1 , that

$$\text{HR}(t) < \phi$$

assuming $\phi < 1$.

- Is $\text{HR}(t)$ decreasing in t ?
- Intuitively one would answer Yes to both of these conjectures.
- Turns out to be true for a certain class of joint distributions of (T^0, T^1) .
- But it is not true in general, see Martinussen et al. (2021).

Remarks

- A HR does not convey a causal effect when interpreted as a HR. Even in a randomized study.
- The Cox hazard ratio does not express, for instance, that treatment works equally effectively at all times, as the hazard ratio at a given time mixes differences between treatment arms due to treatment effect as well as selection.
- Cox model may be used to calculate survival probabilities.
- These have a causal interpretation. We return to this in a moment.
- The problem is not unique to the Cox model.
- Problem is still there for the additive hazard model.
- A hazard ratio that does have a causal interpretation is

$$\text{HR}(t) = \frac{P(T^1 = t | T^0 \geq t, T^1 \geq t)}{P(T^0 = t | T^0 \geq t, T^1 \geq t)}$$

But it can never be estimated from the observed data.

- For Cox, we have

$$\exp(\beta) = \frac{\log P(T^1 > t)}{\log P(T^0 > t)}.$$

so in this respect, we have a causal interpretation.

Remarks

- Time-varying HR's are not causally interpretable. May be used for descriptive purposes when addressing non-causal questions, however.
- Consider time-to death after onset of a certain disease, and say that there are two subclasses ($A = 1$ and $A = 0$) of the disease.
- We know from a previous study where the model change-point model was used, and was correctly specified, that $\exp(\beta_1) = 2$, $\beta_2 = 0$, for $\nu = 2$ years.
- Assume also that all other risk factors were equally balanced in the two subclasses at time 0 (onset of disease) or that proper adjustment for these factors was performed in the hazard model.
- Then, based on this analysis, we can say that patients in disease subclass 1 who survive the first two years from onset of the disease will have the same subsequent risk of dying as those in disease subclass 0
- A statement that may be useful for consulting patients at time $\nu = 2$ years, or later. But not from start of study.

Causal Inference

Causal Inference and the Cox HR

Other measures

Unobserved confounding

IV-methods

Time-to-event data



Other measures to report that has a causal interpretation

Other measures

- The problem with the Cox HR is the conditioning on

$$T \geq t$$

ie the

$$\text{Hazard}(t) = P(T = t \mid T \geq t)/dt$$

Measures that do have a causal interpretation are

$$P(T > t)$$

$$P(T \leq t)$$

Assume first that we have data

$$Z = (T, A, W)$$

A: Treatment (randomised)

W: Covariates

Causal Inference

oooooooooooooooo

Causal Inference and the Cox HR

oooooooooooooooo

Other measures

oo●ooo

Unobserved confounding

oooooooooooo

IV-methods

oooooo

Time-to-event data

oooooooooooooooooooo

Other measures

We assume no unmeasured confounders

$$T^a \perp A = a \mid W$$

and we are interested in comparing

$$P(T^1 \leq t) - P(T^0 \leq t)$$

Note

$$\begin{aligned} P(T^a \leq t) &= E(P(T^a \leq t \mid W)) \\ &= E(P(T^a \leq t \mid A = a, W)) \\ &= E(P(T \leq t \mid A = a, W)) \\ &\equiv \psi_t^a(P) \end{aligned}$$

Based on semiparametric theory an optimal estimating equation for $\psi_t^a(P)$ is given by the efficient influence function

$$\begin{aligned} \tilde{\psi}_t^a(Z) &= \frac{I(A = a)}{P(A = a \mid W)} (I(T \leq t) - P(T \leq t \mid A = a, W)) \\ &\quad + P(T \leq t \mid A = a, W) - \psi_t^a(P) \end{aligned}$$

Other measures

This leads to the estimator

$$\hat{P}(T^a \leq t) = n^{-1} \sum_i [P(T \leq t | A = a, W_i) \\ + \frac{I(A_i = a)}{P(A = a | W_i)} (I(T_i \leq t) - P(T \leq t | A = a, W_i))]$$

To calculate $\hat{P}(T^a \leq t)$ we need working models

- $P(T \leq t | A = a, W)$
- $P(A = a | W)$

Remember

$$P(T \leq t | A = a, W) = 1 - P(T > t | A = a, W)$$

and

$$P(T > t | A = a, w) = \exp \left(- \int_0^t \lambda(s | a, w) ds \right)$$

So here we may use Cox.

Other measures

In survival analysis it is not realistic to assume that the data is

$$Z = (T, A, W)$$

since we will typically have censoring ie data is

$$D = \left(\tilde{T} = T \wedge C, \Delta = I(T \leq C), A, W \right)$$

Let

$$N_C(t) = I(\tilde{T} \leq t)(1 - \Delta)$$

$$Y(t) = I(t \leq \tilde{T})$$

$$dM_C(t) = dN_C(t) - Y(t)d\Lambda_C(t)$$

then the eif based on D is

$$\frac{\Delta \tilde{\psi}_t(Z)}{K_c(T)} + \int \frac{E(\tilde{\psi}_t(Z)|T > s, W)}{K_c(s)} dM_C(s) \quad (4)$$

Where $\tilde{\psi}_t(Z)$ is eif based on the full data Z and

$$K_c(s) = P(C > s)$$

may depend on A, W .

Exercise

We are interested in the parameter $\psi_t^1(P)$. Assume first full data $Z = (T, A, W)$.

- (a) Show that

$$I(T \leq t) - P(T \leq t | A = 1, W) = S(t | A = 1, W) \int_0^t \frac{1}{S(u | A = 1, W)} dM^T(u | A = 1, W)$$

where $S(u | A = 1, W) = P(T > u | A = 1, W)$ and

$$dM^T(u | A = 1, W) = dN^T(u) - I(u \leq T) d\Lambda_T(u | A = 1, W)$$

with $N^T(u) = I(T \leq u)$, and Λ_T is the integrated (conditional) hazard function.

- (b) Rewrite the eif $\tilde{\psi}_t^1(Z)$ based on the full data where you use the result obtained in (a).

Now we consider the censored data setting with $D = (\tilde{T} = T \wedge C, \Delta = I(T \leq C), A, W)$

- (c) Consider formula (4). Show that for any component, say $g(Z)$, in $\tilde{\psi}(Z)$ that is not affected by the censoring (ie observed regardless of censoring) will be unchanged by plugging it into (4), i.e., by plugging it into (4) will just give us $g(Z)$ back.

Assume now that we have n iid replicates $D_i = (\tilde{T}_i = T_i \wedge C_i, \Delta_i = I(T_i \leq C_i), A_i, W_i)$ and are interested in estimating $P(T^1 \leq t)$.

- (d) Derive the eif based on the observed data $\tilde{\psi}_t^1(D)$ (Hint: use formula on previous slide).

- (e) Use Lemma A2 in Lu and Tsiatis (2008) to express $\tilde{\psi}_t^1(D)$ based on the observed martingale $M(t | A = 1, W)$

- (f) Use the result obtained in (f) to propose an estimator for $P(T^1 \leq t)$.

- (g) Comment: Data-example... Use the Cox-Aalen-function to calculate the estimator.

Causal Inference



Causal Inference and the Cox HR



Other measures



Unobserved confounding



IV-methods



Time-to-event data



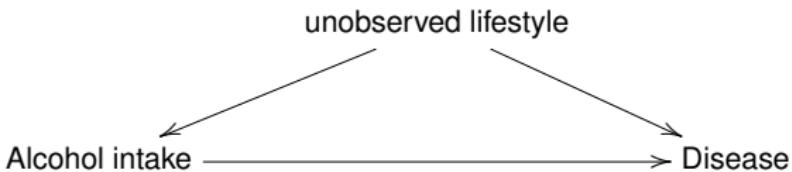
Unobserved confounding and instrumental variables

Motivation

- Can you guarantee that the results from your observational study are unaffected by unmeasured confounding?
- The only answer an epidemiologist can provide is "no".
- Imagine for a moment the existence of an alternative method that allows one to make causal inference from observational studies even if the confounders remain unmeasured.
- That method would be an epidemiologist's dream.
- The above quotes are taken from Hernan and Robins (*Epidemiology*, 2006).

Alcohol consumption

- Can we estimate the exposure effect when there is un-observed confounding?



- Alcohol consumption has been found in observational studies to have positive effects (coronary heart disease) as well as negative effects (liver cirrhosis, some cancers).
- But also strongly associated with all kinds of confounders (lifestyle etc.) as well as subject to self-report bias. Hence doubts in causal meaning of above effects.

Mendelian Randomization: Basic idea

Idea: if we cannot randomise, let's look for instance where nature has randomised, eg through genetic variation.

Example: Alcohol consumption

Genotype: ALDH2 determines blood acetaldehyde, the principal metabolite for alcohol.

Two variants, 1 and 2.

22 homozygotes individuals suffer facial flushing, nausea, drowsiness and headache after alcohol consumption

They hence have low alcohol consumption regardless of other lifestyle behaviours.

IV-idea: check if these individuals have a different risk than others for alcohol related health problems.

Association between alcohol intake and ALDH2 genotype.

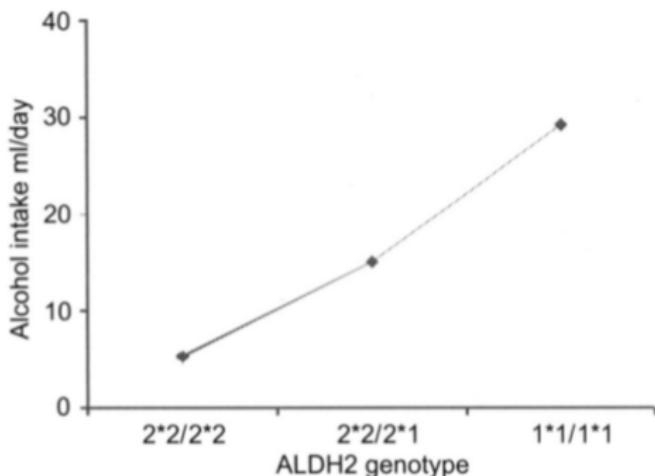
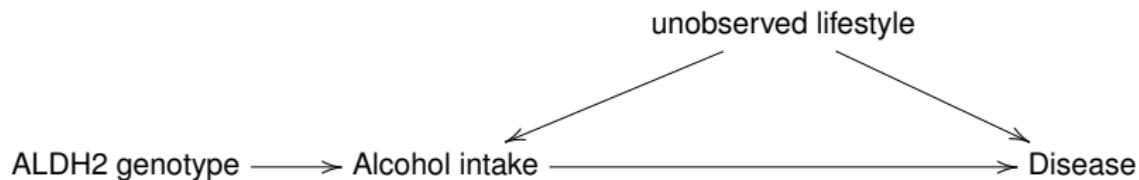


Fig. I. Relationship between alcohol intake and ALDH2 genotype. Data from Takagi et al. (2002).

Although we see an association here, age and cigarette smoking are not related to the genotype! They are of course related to alcohol consumption, however.

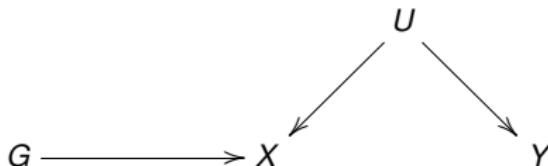
Alcohol consumption



- Causal effect? Under IV-assumptions, the null-hyp of no causal effect of alcohol consumption, should imply no association between ALDH2 and disease.
- While if alcohol consumption has a causal effect we would expect an association between ALDH2 and disease.

Using IV to test for causal effect

- Simply test if Y (outcome) and G (instrument) are independent.



- Regardless of measurement level, testing independence between Y (outcome) and G (instrument) is valid test for presence of causal effect of X (exposure) on Y .

Alcohol consumption

Results (Meta-analysis by Chen et al. (2008)).

- Blood pressure on average 7.44 mmHg higher and risk of hypertension 2.5 higher for ALDH2 11's than 22's carriers (only males).
 - ▶ Mimics the effect of large versus low alcohol consumption.
- Blood pressure on average 4.24 mmHg higher and risk of hypertension 1.7 higher for ALDH2 12's than 22's carriers (only males).
 - ▶ Mimics the effect of moderate versus low alcohol consumption.
 - ▶ indicates that even moderate alcohol consumption is harmful.

Mendelian Randomization

- This was an example of a Mendelian Randomization analysis
- **It is not about** establishing a causal relationship between gene and disease.
- The goal is to investigate whether there is a **causal relationship** between a **modifiable risk factor** (alcohol consumption) and the **disease**.
- Mendelian Randomization analysis is an example of what we more generally call instrumental variables analysis.

Instrumental variables DAG

- Can we estimate the exposure effect when there is un-observed confounding?

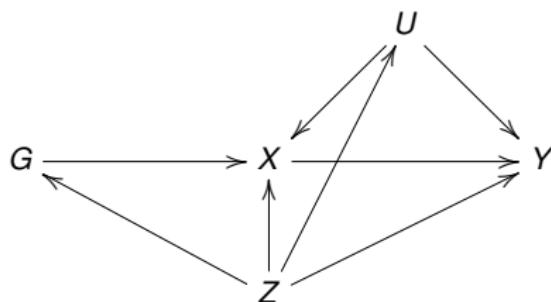
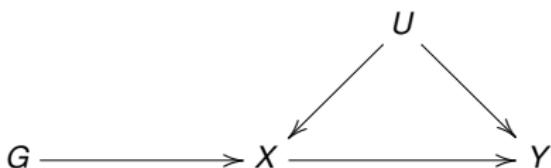


Figure: G is the instrument, X the exposure variable. U : potential unmeasured confounders; observed confounders by Z .

Instrumental variables DAG

Let's look at DAG without additional observed covariates Z :



Core assumptions:

1. G and X are associated
2. G is independent of the unmeasured confounder(s) U
3. G is independent of the outcome given X and U . (Exclusion restriction).

Ass. 2. and 3. are hard (often impossible) to check from the observed data.

OBS: The 3 ass. above are actually not enough to do estimation!

Examples of instruments

- Genes; Mendelian randomization.
- Treatment assignment in randomized clinical trials with non-compliance.
- Physician treatment prescribing preference.
- Martens et al. (Epi, 2006).

Using IV to estimate causal parameters

- Target parameters (continuous response; binary response; time-to-event response).
- Classical 2-stage-least square estimator.
- Methods for time-to-event outcome data
- Methods for competing risk data.

Causal Inference

oooooooooooooooooooo

Causal Inference and the Cox HR

oooooooooooooooooooo

Other measures

ooooooo

Unobserved confounding

ooooooooooooooo

IV-methods

o●oooo

Time-to-event data

oooooooooooooooooooo

Continuous outcome Y

We will assume the (underlying true) model

$$E(Y|X, G, Z, U) = \beta_0 + \beta_X X + \beta_Z Z + \beta_U U$$

If we had information about the confounder(s) U we could just do ordinary regression

```
> n=10000
> U=rnorm(n)
> G=rbinom(n,1,0.5)
> Z=rbinom(n,1,0.5)
> alp0=1;alp1=0.75;alp2=-0.5;alp3=0.3
> mu=alp0+alp1*G+alp2*U+alp3*Z
> X=rnorm(n,mu,1)
>
> bet0=1;bet1=0.5;bet2=-1;bet3=0.3
> thet=bet0+bet1*X+bet2*U+bet3*Z
> Y=rnorm(n,thet,0.5)
```

Hence, the β_X is 0.5 in this scenario. Can we estimate it from data?

Continuous outcome Y

If we had information about the confounder(s) U we could just do ordinary regression:

```
> summary(lm(Y~X+Z+U))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.988966  0.009615 102.86 <2e-16 ***
X           0.503275  0.004709 106.88 <2e-16 ***
Z           0.321210  0.010139  31.68 <2e-16 ***
U          -0.997307  0.005549 -179.72 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem is that we don't get to see U !

So we cannot just do the above regression.

Continuous outcome Y

Do the regression without U , that is, use the data we actually get to see (so this may be the problem with an observational study):

```
> summary(lm(Y~X+G+Z))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.575063  0.019928  28.86 <2e-16 ***
X            0.894512  0.009174  97.51 <2e-16 ***
G            -0.280218  0.021542 -13.01 <2e-16 ***
Z            0.220605  0.020671  10.67 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem is that we now get a biased estimate of effect of X ; remember true coefficient was 0.5.

Note also that we see a significant effect of G in the above analysis. But we know that there should be no such one. Can you explain why this happens? Use the underlying DAG.

Continuous outcome Y

Is there a causal effect of X ?

```
> summary(lm(Y~G))  
Coefficients:  
            Estimate Std. Error t value Pr(>|t|)  
(Intercept) 1.72321   0.02038  84.56 <2e-16 ***  
G            0.37182   0.02903  12.81 <2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Can we somehow estimate the correct effect of X without using U (which we don't observe)?
- Yes, we can! Under the core assumptions (and actually a little more).
- It is called 2SLS estimation
- In this scenario the so-called Wald estimator also gives the same.

Continuous outcome Y

So what are these estimates?

The 2SLS estimator:

- Do the regression of X on G (and Z), and calculated the predicted values: \hat{X}
- Do the regression of Y on \hat{X} (and Z)

```
> hat.X=fitted(lm(X~G+Z))
> summary(lm(Y~hat.X+Z))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.96376   0.05743   16.78   <2e-16 ***
hat.X        0.50903   0.03936   12.93   <2e-16 ***
Z            0.33775   0.03105   10.88   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So, we get the estimate 0.509.

The above reported s.e. is incorrect, but a correct one is available.

Time-to-event data

- Assume the outcome is a (possibly right-censored) time-to-event outcome
- Example: time-to-death; exposure vitamin D.
- Not so much work concerning IV analysis for this type of outcome has been done.
- This has been within my own research area.
- Papers: Tchetgen Tchetgen et al. (Epidemiology, 2015); Martinussen et al. (Biometrics, 2017).
- Both use additive hazards models (absolute risk).
- In the first one we develop a 2SLS method, but it is based on stronger assumptions than the last one.
- In Martinussen et al. (Biostatistics, 2017), we develop a method using an IV in Cox-regression setup.

Vitamin and mortality. Some more details.

- Vitamin D deficiency is common and has been linked with a number of common diseases such as cardiovascular disease, diabetes, and cancer.
- Here we will be interested in evaluating the causal effect on overall survival.
- Unfortunately, vitamin D status is also associated with several behavioral and environmental factors potentially giving rise to biased estimators when using standard statistical analyses.
- Recently, mutations in the filaggrin gene have been shown to be associated with a higher vitamin D status, supposedly through an increased UV sensitivity of keratinocytes (see Skaaby et al., 2013, and references therein).
- To assess the potential causal effect of vitamin D status on overall survival, we used filaggrin genotype as a proxy measure for vitamin D status according to the principles of Mendelian randomisation.
- The ten-year follow-up study, Monica10, included 2656 individuals between 40-71 years. It was carried out in 1993-1994,

Time-to-event response

- Assume to begin with that:

$$\lambda(t; X, G, U) = \beta_0(t) + \beta_X(t)X + \beta_G(t)G + \beta_U(t)U$$

with $\beta_G(t) = 0$.

- In the following I will describe two different approaches, one is the 2SLS
- Counting process notation:

$$N(t) = I(T \leq t) \quad R(t) = I(t \leq T)$$

- So

$$E\{N(t)\} = E\left\{\int_0^t R(s)\lambda(s) ds\right\}.$$

- This also easily handles (independent) right-censoring.

2SLS for survival outcome

- Still assume Aalen's additive hazards model.
- Assume now also that $X = c_0 + c_G G + \Delta$; $E(\Delta|G) = 0$.
- It is now easy to see that

$$\begin{aligned} P(T > t|G) &= E(e^{-B_0(t)-B_X(t)X-B_U(t)U}|G) \\ &= E(e^{-B_0(t)-B_X(t)(c_0+c_GG+\Delta)-B_U(t)U}|G) \\ &= e^{-\tilde{B}_0(t)-\textcolor{blue}{B_X(t)}V}, \end{aligned}$$

where $V = c_0 + c_G G$

- Let $c = (c_0, c_G)$, $\hat{V}_i = \hat{c}(1, G_i)^T$ and $\hat{V}(t)$ be the $n \times 2$ -matrix with i th row $(R_i(t), R_i(t)\hat{V}_i)$.
- The causal exposure effect is therefore estimated by

$$\hat{B}(t) = \int_0^t \{\hat{V}(t)^T \hat{V}(t)\}^{-1} \hat{V}(t)^T dN(t),$$

- Easy to work out large sample properties. So this works for continuous X .
- R-package (not yet on Cran though), `ivtools` (Sjølander, Dahlqwist and Martinussen) that can do all the mentioned estimation. source-code on homepage.

Causal effect of Vitamin D

- Vit D (nmol/l)

```
> install.packages("ivtools")
> library("ivtools")
> data(VitD)
> head(VitD)
  age filaggrin vitd      time death    n_vitd
1  41          0 53.3 16.16220     0 -0.4207751
2  62          0 26.4 16.06470     0 -1.4157752
3  41          0 47.7 16.34796     0 -0.6279127
4  42          0 103.5 16.32993    0  1.4360651
5  52          0 79.0 16.74289    0  0.5298383
6  62          0 70.1 15.17891    0  0.2006375

> table(VitD$filaggrin)
  0   1
2377 194

> summary(lm(vitd~filaggrin,data=VitD))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 64.2633    0.5538 116.034 < 2e-16 ***
filaggrin    5.4656    2.0162   2.711  0.00676 ** 
---
> VitD$n_vitd=(VitD$vitd-mean(VitD$vitd))/sd(VitD$vitd) # Normalize
```

- Have also information on age. Need to adjust for that.

Causal effect of Vitamin D?

```
> fit.aalen=aalen(Surv(time,death)~age+filaggrin,max.time=10,data=VitD)
> summary(fit.aalen)
```

Additive Aalen Model

Test for nonparametric terms

Test for non-significant effects

	Supremum-test of significance p-value H_0: B(t)=0
(Intercept)	11.70 0.000
age	13.20 0.000
filaggrin	2.69 0.075

Test for time invariant effects

	Kolmogorov-Smirnov test p-value H_0:constant effect
--	---

(Intercept)	0.10300 0.000
age	0.00229 0.000
filaggrin	0.01180 0.902

##

Parametric terms :

	Coef.	SE	Robust SE	z	P-val	lower2.5%	upper97.5%
const(filaggrin)	-0.00276	0.00251	0.00255	-1.08	0.279	-0.00768	0.00216

Causal effect of Vitamin D. 2SLS-approach

```
> ## 2SLS
> fitX.LZ=glm(n_vitd~filaggrin+age,data=VitD)
> fitT.LX=ah(Surv(time,death)~ n_vitd+age,data=VitD)
> fit.IV1=ivah(estmethod="ts",fitX.LZ=fitX.LZ,fitT.LX=fitT.LX,data=VitD,ctrl
> summary(fit.IV1)

Call:
ivah(estmethod = "ts", fitX.LZ = fitX.LZ, fitT.LX = fitT.LX,
     data = VitD, ctrl = F)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
n_vitd -0.026080   0.014767 -1.766   0.0774 .
age      0.001360   0.000105 12.946 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Time-to-event response. G-estimation

Another and more general approach for which we center G so that $E(G) = 0$. (Replace G_i with $G_i - \bar{G}$).

Estimation

$$E\{Ge^{B_X(t)X} R(t)(dN(t) - dB_X(t)X)\} = 0$$

- Proof:

$$\begin{aligned} LHS &= EE(|\mathcal{F}_t^{X,G,U}) = E\{Ge^{B_X(t)X} R(t)(dB_0(t) + dB_U(t)U)\} \\ &= EE(|X, G, U) \\ &= E\{Ge^{B_X(t)X} e^{-B_0(t)-B_X(t)X-B_U(t)U}(dB_0(t) + dB_U(t)U)\} \\ &= E\{G \cdot h(t, U)\} = 0 \end{aligned}$$

Time-to-event response

Estimation

$$E\{Ge^{B_X(t)X}R(t)(dN(t) - dB_X(t)X)\} = 0$$

- Estimator:

$$\hat{B}_X(t) = \int_0^t \frac{\sum_i G_i e^{\hat{B}_X(s-)X_i} dN_i(s)}{\sum_i G_i R_i(s) e^{\hat{B}_X(s-)X_i} X_i}$$

with $G_i = G_i - \bar{G}$.

- Note the recursive structure of $\hat{B}_X(t)$.
- Explain how to carry this out in practice (5 minutes).

Time-to-event response

- Can reformulate in a slightly more general way:
- We focus on the so-called local causal effect of exposure X :

$$\frac{P(T^x > t | X = x, G)}{P(T^0 > t | X = x, G)} = \frac{P(T > t | X = x, G)}{P(T^0 > t | X = x, G)} = e^{-B_X(t)x}$$

- $B_X(t)$ is the effect of removing exposure for those who are exposed and keeping G fixed.
- It still holds that

$$(*) = E\{Ge^{B_X(t)X} R(t)(dN(t) - dB_X(t)X)\} = 0$$

and therefore the same estimator applies!

Asymptotics

- Estimator: $\hat{B}_X(t) = \int_0^t \frac{\sum_i G_i e^{\hat{B}_X(s-) X_i} dN_i(s)}{\sum_i G_i R_i(s) e^{\hat{B}_X(s-) X_i} X_i}$ with $G_i = G_i - \bar{G}$.
- Note the recursive structure of $\hat{B}_X(t)$.
- Actually, $\hat{B}_X(t) = \hat{B}_X(t, \hat{\theta})$
- where $\theta = E(G)$ and $\hat{\theta} = \bar{G}$
- And this is also extended to handle additional observed confounders L :
 $\mu(L; \theta) = E(G|L; \theta)$
- One may show that $\hat{B}_X(t)$ is consistent, not so easy to show.

Asymptotics

- For known θ :

$$\hat{B}_X(t, \theta) = \int_0^t H\{s, \hat{B}_X(s-, \theta)\} dN(s),$$

where the k th element of the n -vector $H\{s, \hat{B}_X(t-, \theta)\}$ is

$$\{G_k - \mu(L_k; \theta)\} e^{\hat{B}_X(t-, \theta) X_k} / \sum_i \{G_i - \mu(L_i; \theta)\} e^{\hat{B}_X(t-, \theta) X_i} R_i(t).$$

- Let $V(t, \theta) = n^{1/2}\{\hat{B}_X(t, \theta) - B_X(t)\}$ and \dot{H} the derivative of H with respect to second argument.
- $V(t, \theta)$ solves a Volterra-equation:

$$\begin{aligned} V(t, \theta) &= n^{1/2} \int_0^t H(s, B_X(s-)) [dN(s) - X dB_X(s)] \\ &\quad + \int_0^t V(s-, \theta) \dot{H}(s, B_X(s)) dN(s) + o_p(1) \end{aligned}$$

Asymptotics

- The solution to this equation is given by

$$V(t, \theta) = \int_0^t \mathcal{F}(s, t) n^{1/2} H(s, B_X(s-)) [dN(s) - X dB_X(s)],$$

where

$$\mathcal{F}(s, t) = \prod_{(s, t]} \left\{ 1 + \dot{H}(\cdot, B_X(\cdot)) dN(\cdot) \right\}$$

- This leads to the iid-representation

$$V(t, \theta) = n^{-1/2} \sum_{i=1}^n \epsilon_i^B(t)$$

with the $\epsilon_i^B(t)$'s being zero-mean iid terms. Specifically

$$\epsilon_i^B(t) = \int_0^t \mathcal{F}(s, t) n^{1/2} \{H(s, B_X(s-))\}_i [dN(s) - X dB_X(s)]_i$$

Causal Inference

oooooooooooooooo

Causal Inference and the Cox HR

oooooooooooooooo

Other measures

oooooo

Unobserved confounding

oooooooooooo

IV-methods

oooooo

Time-to-event data

oooooooooooo●oooooooo

Causal effect of Vitamin d

```
> fitZ.L=glm(filaggrin~age,family="binomial",data=VitD) # fit of instrument on covariates
> fit.IV2=ivah(estmethod="g",X="n_vitd",T="time", fitZ.L=fitZ.L,event="death",data=VitD,n=1000)
> summary(fit.IV2)

Test for non-significant exposure effect. H_0: B(t)=0

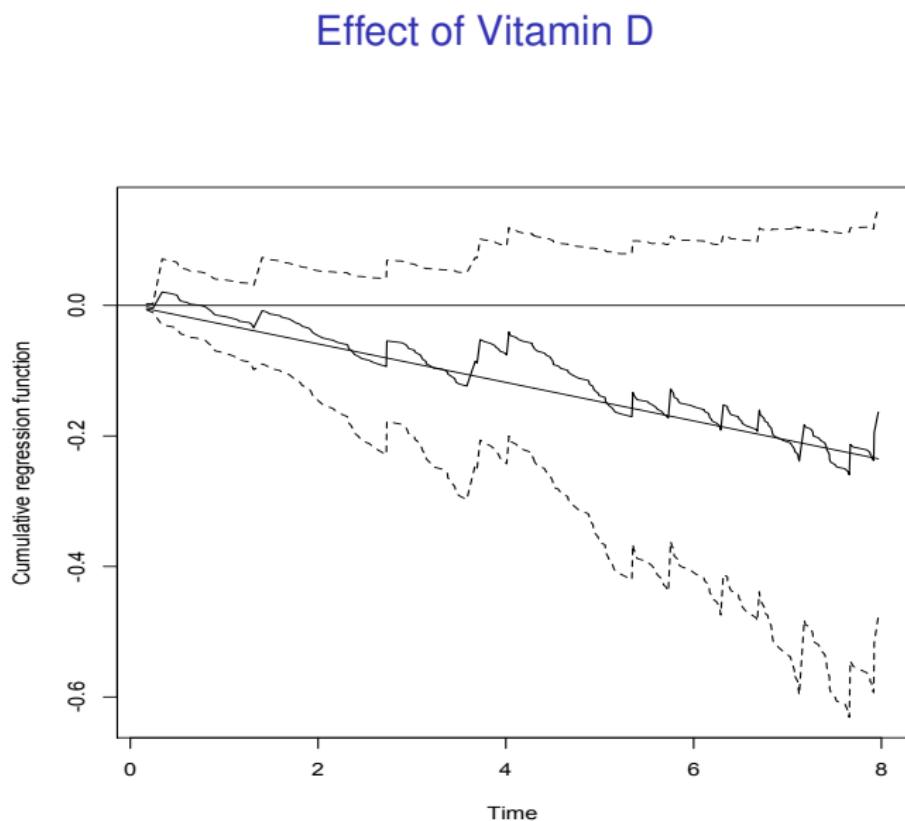
Supremum-test pval
n_vitd          0.3

Goodness-of-fit test for constant effects model
Supremum-test pval
                0.96

Constant effect model

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
n_vitd -0.02212   0.01917  -1.154   0.248

> plot(fit.IV2)
> abline(0,0)
```



LIVER data from ECOG-trial E9288. Exercise

- The data are based on 109 individuals with colorectal cancer, who were scheduled for surgery resection due to liver metastases
- Purpose was to evaluate a new treatment, whereby the patient while undergoing surgery was implanted with an arterial device (AD) to administer chemotherapy.
- 56 patients were randomized to surgical resection alone (standard treatment), and 53 patients were randomized to the new treatment intervention thus also receiving chemotherapy.
- Perfect compliance in the standard group because the new treatment was only available for those randomized to it, there was non-compliance in the intervention arm. In this arm, 10 patients did not receive the AD for unreported reasons, presumably related to their prognosis.
- Can we use an IV-analysis? Explain setting, why not standard analysis, what is the instrument?
- Get the data:

```
>load(file=url("http://publicifsv.sund.ku.dk/~tma/CausalInf/2018/Liver_dat.Rda"))
>head(Liver_dat)
```

- Do the IV additive hazards analysis for these data (`ivah`) - use the G-estimation approach.

IV-estimation within the Cox-model

- Consider now

$$\frac{\lambda_{T^x}(t|X=x, G)}{\lambda_{T^0}(t|X=x, G)} = e^{\beta x}$$

- Or, with $g(v) = \log(-\log(v))$,

$$g(P(T^x > t_0 | X = x, G)) = g(P(T^0 > t_0 | X = x, G)) + \beta x$$

- Estimation may proceed exploiting that Y^0 and G are independent. If G and X are not both categorical then we need a model for either $P(T^x > t_0 | X = x, G)$ or $P(T^0 > t_0 | X = x, G)$.
- Assume Cox:

$$\lambda_T(t|X=x, G=g) = \lambda_0(t)e^{\phi_X x + \phi_G g}$$

IV-estimation within the Cox-model

- Assume for the moment $E(G) = 0$. Since

$$\begin{aligned} 0 &= E(GP(T^0 > t)) = E(GP(T^0 > t|G)) = E(GP(T^0 > t|X, G)) = \\ &= E(G \exp(-\Lambda_0(t)e^{\phi_X X + \phi_G G} e^{-\beta X})) \end{aligned}$$

- An estimating equation is then given by $0 = U(\beta)$, where

$$U(\beta) = \sum_i (G_i - \bar{G}) \exp(-\hat{\Lambda}_0(t_0)e^{\hat{\phi}_X X_i + \hat{\phi}_G G_i} e^{-\beta X_i})$$

- May give some problems concerning so-called incongeniality.

References

- 2SLS** Instrumental variable estimation in a survival context. Eric J. Tchetgen Tchetgen, Stefan Walter, Stijn Vansteelandt, Torben Martinussen, Maria Glymour. *Epidemiology*, 2015.
- Aalen** Instrumental variables estimation of exposure effects on a time-to-event response using structural cumulative failure time models. Torben Martinussen, Stijn Vansteelandt, Eric J. Tchetgen Tchetgen and David M. Zucker. *Biometrics*, 2017.
- Cox** Instrumental variable estimation for survival data using a structural nested Cox model. Torben Martinussen, Ditte Nørbo Sørensen and Stijn Vansteelandt. *Biostatistics*, 2017.
- Software** There is an R-package, `ivtools` (Sjølander, Dahlqwist and Martinussen) that can do all the mentioned estimation.

LIVER data from ECOG-trial E9288.

- The data are based on 109 individuals with colorectal cancer, who were scheduled for surgery resection due to liver metastases
- Purpose was to evaluate a new treatment, whereby the patient while undergoing surgery was implanted with an arterial device (AD) to administer chemotherapy.
- 56 patients were randomized to surgical resection alone (standard treatment), and 53 patients were randomized to the new treatment intervention thus also receiving chemotherapy.
- Perfect compliance in the standard group because the new treatment was only available for those randomized to it, there was non-compliance in the intervention arm. In this arm, 10 patients did not receive the AD for unreported reasons, presumably related to their prognosis.
- This model can also be fitted using the function `ivcoxph` in the R-package `ivtools`, see pages 12-15 in manual for `ivtools` (google).
- You may now use `ivcoxph` to analyse the LIVER data from ECOG-trial E9288.
- You may want to compare results with a standard Cox-analysis

Exercise on model congeniality

Let T be a time-to-event variabel, X and G are exposure and instrument, respectively. We shall assume that

$$T^0 | X, G \sim \lambda_0(t) e^{\tilde{\beta}X + \tilde{\gamma}G},$$

ie a Cox model. We further assume that the DGP is as follows: G is binomial(1,p), $X = \alpha_0 + \alpha_1 G + \epsilon_X$ with $\alpha_1 \neq 0$ and $\epsilon_X \sim N(0, 1)$ (and ϵ_X independent of anything), and

$$T | X, G \sim \lambda_0(t) e^{\beta X + \tilde{\gamma}G}, \quad (5)$$

with $\beta = \tilde{\beta} + \psi$. We are interested in the structural Cox model

$$\lambda_{Tx}(t | X = x, G) = \lambda_{T^0}(t | X = x, G) e^{\psi x} \quad (6)$$

- (a) For G to be a valid instrument we need to have G and T^0 independent. Show that this restricts the parameters as follows

$$\beta = \psi - \gamma / \alpha_1. \quad (7)$$

In this case we say that models (1) and (2) are congenial, because there is in fact a DGP so that both models can be correctly specified, and so that the IV-assumptions hold.

The association model (model for the observed data) and the structural model are said to be incongenial if there is no DGP so that both models are correctly specified.

Now assume we have iid observations from the above DGP where (3) holds.

- (b) Exploit the independence between G and T^0 to suggest an estimating equation for the parameter of interest, ψ .

Exercise on model congeniality (Ctd)

To avoid the issue of incongeniality one may proceed as follows. We still have interest in model (2), but we are not going to model the observed data (no association model), except that we still assume that G is binomial(1,p) and $X = \alpha_0 + \alpha_1 G + \epsilon_X$ with $\alpha_1 \neq 0$, $\epsilon_X \sim N(0, 1)$. Define now the so-called selection bias function

$$q(t, X, G) = -\log \left\{ \frac{S_{T^0}(t|X, G)}{S_{T^0}(t|X = 0, G)} \right\}$$

with $q(t, 0, G) = 0$. The selection function q can take any real value; a negative value expresses that $S_{T^0}(t|X, G)$ is larger than the reference survival had the individuals in this strata been unexposed. This means that due to some selection mechanism, the exposed individuals have a better survival than the unexposed, had we been able to remove the exposure.

- (c) Show that T^0 and G are independent if and only if

$$S_{T^0}(t|X = x, G) = \exp \left[-q(t, x, G) - \log E \left\{ e^{-q(t, X, G)} | G \right\} - \Omega(t) \right],$$

where

$$\Omega(t) = -\log \{ S_{T^0}(t) \},$$

Assume that

$$q(t, x, G) = B_X(t)x,$$

which is referred to as homogen selection bias (so instead of specifying an association model, and running the risk of model in-congeniality, we model the selection bias function). In this way we ensure that $T^0 \perp\!\!\!\perp G$ and at the same time avoid model in-congeniality.

Assume that $\Omega(t) = \int_0^t \omega(s)ds$ and $B_X(t) = \int_0^t \beta_X(s)sds$ (ie that the two function are absolutely continuos).

- (d) Find the observed hazard function, i.e., find $\lambda_T(t|X, G)$, and conclude that is on the so-called Cox-Aalen model form. Hint: Remember, $\lambda(t) = -D_t \log S(t)$.

Unmeasured confounding and competing risks

- Time-to-event analyses are often plagued by both – possibly **unmeasured – confounding and competing risks.**
- HIP trial on effectiveness of screening on breast cancer mortality:
 - ▶ About 60000 women aged 40-60 were randomized into two equally sized groups.
 - ▶ Approximately 35% of the women in the screening group refused to participate (non-compliers) There were large differences between the study women who participated and those who refused (Shapiro, 1977)
 - ▶ Results from the "as treated" analysis may be doubtful due to unobserved confounding.
 - ▶ There is further a competing risk issue in these data. In the first 10 years of follow-up there are 4221 deaths but only 340 were deemed due to breast cancer.
 - ▶ An IV-analysis is tempting as the original randomization can be used as instrument.
 - ▶ How can we do this when dealing with competing risk data?
 - ▶ Martinussen and Vansteelandt, Biostatistics 2018.