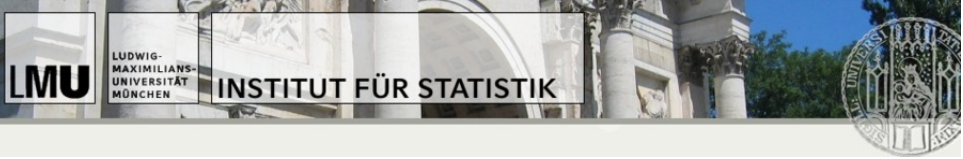


Statistical Methods in Epidemiology

Chapter 9: Time-to-Event Data & Competing Risks

Winter Semester 2022/23
Version: January 24, 2023

Michael Schomaker
Ludwig-Maximilians-Universität München, Institut für Statistik

What is time-to-event data?

Data where the outcome consists of the tuple (T, Y):

Time T: duration from a predefined time zero until an event of interest occurs

Event Y: for example, death, clinical event, viral failure, recovery

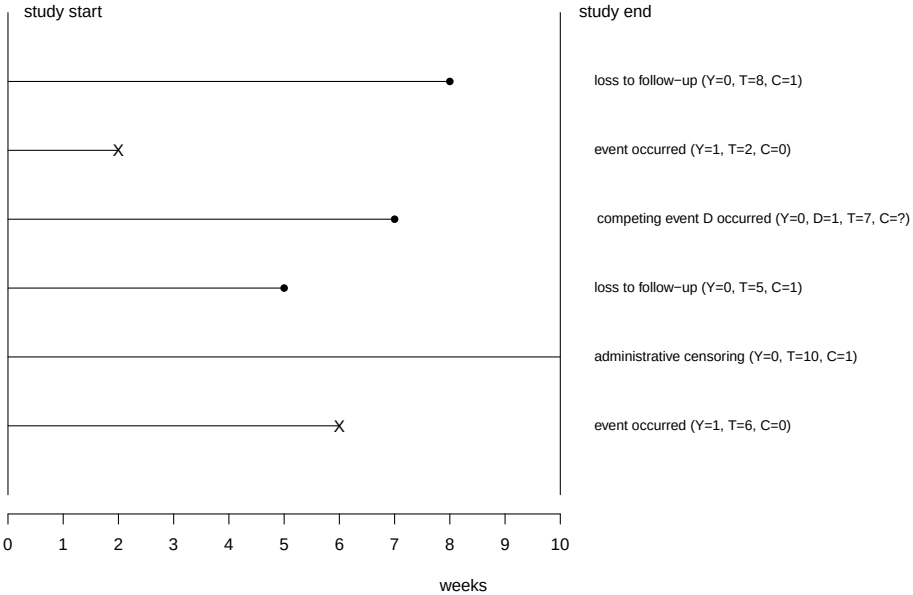
Censoring (I)

Almost always, for many subjects we do not observe the event/outcome due to

- study end at a particular date
- participants leave the study (drop out, loss to follow-up)
- another, competing event D occurs
- patients get transferred to other facilities
- patients leave insurance schemes / database

Units for which any of the above applies get “censored” at the respective date and $Y = 0$ at this date (we may define $C = 1$).

Censoring (II)



Example

We consider the data example from Young et al. [1]:

- T = time to death due to prostate cancer
- Y = death due to prostate cancer
- D = death due to other causes
- C = censoring due to end of follow-up
- A = high dose estrogen therapy (A=1) versus placebo (A=0)
- L = age, blood pressure, clinical stage etc.

Example (II)

	T	event_type	Y	D	A	age	weight_index	stage	normal_activity	hemoglobin	SBP
3	40	odeath	0	1	1	69	102	3	TRUE	1	14
5	65	alive	0	0	0	67	99	3	TRUE	0	17
7	46	odeath	0	1	0	75	100	3	TRUE	0	14
8	62	alive	0	0	0	73	114	3	TRUE	1	17
12	59	alive	0	0	1	74	105	3	TRUE	1	18
14	49	pdeath	1	0	0	55	112	3	TRUE	1	16
15	20	odeath	0	1	1	73	88	3	TRUE	0	19
16	3	odeath	0	1	1	87	81	3	FALSE	1	17
17	58	alive	0	0	0	64	90	3	TRUE	0	14
19	26	odeath	0	1	1	62	90	3	TRUE	1	13
20	52	alive	0	0	1	74	107	3	TRUE	0	13

If we define follow-up time to be 60 weeks, then we also have to censor patients if they are alive at 60 weeks ($T = 60$, $C = 1$, $Y = 0$, $D = 0$)

Example (III)

We can also work with discrete time / time intervals:

hemoglobin	SBP	DBP	metastases	A	time0	C	D	Y	id
13.398438	14	8	0	1	4	0	0	0	1
13.398438	14	8	0	1	8	0	0	0	1
13.398438	14	8	0	1	12	0	0	0	1
13.398438	14	8	0	1	16	0	0	0	1
13.398438	14	8	0	1	20	0	0	0	1
13.398438	14	8	0	1	24	0	0	0	1
13.398438	14	8	0	1	28	0	0	0	1
13.398438	14	8	0	1	32	0	0	0	1
13.398438	14	8	0	1	36	0	0	0	1
13.398438	14	8	0	1	40	0	1	*NA*	1
13.398438	14	8	0	1	44	*NA*	*NA*	*NA*	1

Notation:

$L_t =$ Covariates at time t

$A_t =$ Intervention at time t

$Y_t =$ Outcome / Event at time t

$C_t =$ Censoring indicator at time t

$D_t =$ Indicator for other, competing event at time t

Example (IV)

In our example, only Y, C, and D vary; but not A and L:

age	stage	norm_a	hx	hemoglobin	SBP	DBP	metastases	A	C.4	D.4	Y.4
87	3	0	1	13.398438	17	12	0	1	0	1	NA
76	4	1	1	8.199219	16	6	1	1	0	0	0
70	4	0	0	7.799805	12	8	1	0	0	0	1
55	4	1	0	14.798828	13	9	1	1	0	0	0
80	4	0	1	11.699219	14	8	0	1	0	1	NA
67	4	1	0	9.599609	12	9	1	1	0	0	0

C.8	D.8	Y.8	C.12	D.12	Y.12	C.16	D.16	Y.16	C.20	D.20	Y.20
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0	0	1	NA	NA	1	NA	NA	1	NA	NA	1
NA	NA	1	NA	NA	1	NA	NA	1	NA	NA	1
0	0	0	0	0	0	0	0	0	0	0	0
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
0	0	1	NA	NA	1	NA	NA	1	NA	NA	1

Implications

In the discrete time-to-event setup, with wide format, we have to consider the following:

- We have to choose a time-ordering of variables in each interval
- If $Y_t = 1$, then $Y_{t+1} = 1$ by definition: "once you are dead, you stay dead", i.e. the probability of an event should not decrease over time
- There are so-called recurrent event setups for which this does not apply, but we do not consider them here
- If $D_t = 1$, then $Y_{t+1} = 0$ (NA) by definition: individuals who experience a competing event can never experience the event of interest anymore; also $Y_{t-1} = 0$.

Estimands without competing risks: the risk

For now, we assume that A does not vary over time.

The *risk* of the event of interest at time t, had all individuals been assigned $A = a$ is:

$$P(Y_t^a = 1)\,.$$

For a binary treatment, we can define the ATE as:

$$P(Y_t^1 = 1) - P(Y_t^0 = 1)\,.$$

The discrete-time *hazard* of the event of interest at time t under $A = a$ is:

$$P(Y_t^a = 1 | Y_{t-1}^a = 0)$$

This is identical to

$$P(T^a \in (t-1, t] | T^a > t-1)$$

where T^a is the counterfactual time to the event under a.

The continuous-time *hazard* of the event of interest at time t under $A = a$ is:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} P(T^a \in (t, t + \Delta t] \mid T^a > t)$$

Without intervening on A, we have the observed data hazard:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t < T \leq t + \Delta t] \mid T > t) \tag{1}$$

Standard time-to-event analysis is, for example, concerned with the PDF [f(t)], CDF [F(t)] and survival function [1-F(t)] of T; not T^a.

More on the hazard

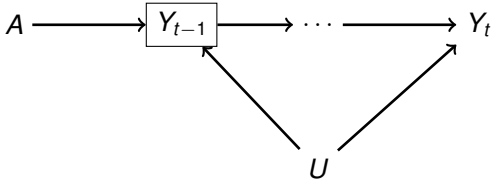
- Unlike the risk, the hazard is defined *conditional* on survival until time $t-1$
- If

$$P(Y_t^1 = 1 | Y_{t-1}^1 = 0) \neq P(Y_t^0 = 1 | Y_{t-1}^0 = 0),$$

  this still does *not* necessarily imply that A has an effect on Y!
- This is because the hazards t may differ simply because of different individuals who survive until $t-1$ under $a=1$ versus $a=0$ due to treatment effects before time $t-1$

More on the hazard (II)

The graphical intuition on why the difference or ratio of two hazards can typically not be interpreted as a causal effect, is selection (collider) bias:



We will discuss this important point later in detail, under the “hazard of hazard ratios” section

Notation: we use the overbar to denote the history of a variable:

$\bar{Y}_t = (Y_1, \ldots, Y_t)$

$\bar{D}_t = (D_1, \ldots, D_t)$

$\vdots$

The *risk* under elimination of competing events is defined as

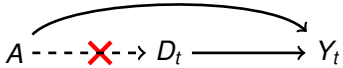
$$P(Y_t^{a,\bar{d}_t=0} = 1)$$

and also known as *marginal cumulative incidence* or *net risk*.

Example: In the prostate cancer example, we can ask: "What is the risk of having died due to prostate cancer at 60 weeks, if it was not possible to have died for any other reason".

Risk under elimination of competing event

- Intervening on the competing event may not always be practically feasible or meaningful (e.g., death); thus, it is sometimes recommended to avoid this estimand for such competing events.
- For some competing events (e.g., D as drop out) the interpretation may be meaningful; and we will come back to this point later.
- Important consideration: the competing event may likely act as a *mediator* between A and Y_t:



Intervening on D_t removes the arrow and thus the indirect effect of A and Y_t through D_t.

- It follows that the risk under elimination of the competing event relates to a *direct* effect.
- Note that we need the arrow $D_t \to Y_t$ because, by definition, $Y_t = 0$ if $D_t = 1$.
- The average treatment effect

$$P(Y_t^{1,\bar{d}_t=0} = 1) - P(Y_t^{0,\bar{d}_t=0} = 1) \tag{2}$$

  is called the *controlled direct effect* and a possible effect measure under competing events.

Risks without elimination of the competing event

The risk without elimination of competing events is

$$P(Y_t^a = 1)$$

and also called the *cause-specific cumulative incidence, crude risk,* or *subdistribution function*. It can be represented as

$$P(T^a < t, \mathcal{D}^a = 1)$$

where $\mathcal{D}^a = 1$ refers to the event of interest and $\mathcal{D}^a = 2$ to the competing event (under $A = a$).

Similarly, we can define the risk of the competing event as

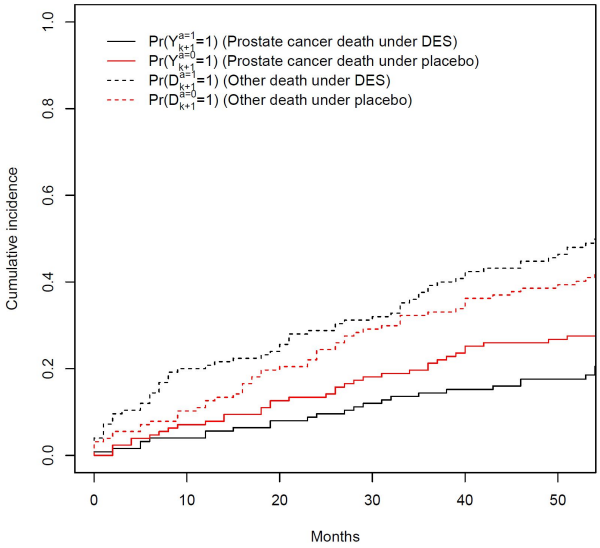
$$P(D_t^a = 1)$$

which is the *cause-specific cumulative incidence, crude risk,* or *subdistribution function* for cause $\mathcal{D}^a = 2$.

It makes sense to present both risks simultaneously as they are deterministically related, i.e. the risk of $P(Y_t^a = 1)$ depends on how many competing events occurred under a before time t, and vice versa.

Example: We can ask what the risk of death due prostate cancer would have been at time t, under either estrogen therapy and placebo, if we did not intervene on competing reasons of death; but calculate those risks too.

Prostate cancer example[1]



[1] Source: Stensrud et al. [2]. We'll discuss estimation of these curves later

Based on the above considerations, we can define the following average treatment type of causal effect measures:

$$P(Y_t^1 = 1) - P(Y_t^0 = 1) \tag{3}$$
$$P(D_t^1 = 1) - P(D_t^0 = 1) \tag{4}$$

and

$$P(Y_t^1 = 1)/P(Y_t^0 = 1)$$
$$P(D_t^1 = 1)/P(D_t^0 = 1)$$

A simple estimand in the presence of competing events is the result of redefining the event of interest as a composite outcome of both the event of interest and the competing event(s)

$$P(Y_t^a = 1 \text{ or } D_t^a = 1),$$

with the corresponding effect measure:

$$P(Y_t^1 = 1 \text{ or } D_t^1 = 1) - P(Y_t^0 = 1 \text{ or } D_t^0 = 1) \tag{5}$$

Example: We can ask for the all-cause probability of death under either estrogen therapy or the placebo. In the given example, this may however not make sense as the whole point of the therapy is to evaluate whether it is helpful with respect to prostate cancer.

Conditional risks

We can evaluate the risk among those individuals who did not experience any competing event.

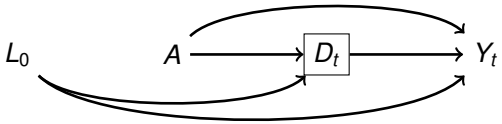
$$P(Y_t^a = 1 \mid D_t^a = 0)\,,$$

A corresponding effect measure would be:

$$P(Y_t^1 = 1 \mid D_t^1 = 0) - P(Y_t^0 = 1 \mid D_t^0 = 0) \tag{6}$$

Intuitively clear that the number of individuals with competing events may be different under $a = 0$ and $a = 1$; thus, such effect measures may not be meaningful.

Additionally, conditioning on D_t may block the indirect effect and add collider bias:



The *Survivor Average Causal Effect (SACE)*, or *Principal Stratum Effect*, is

$$P(Y_t^1 = 1 | D_t^0 = 0, D_t^1 = 0) - P(Y_t^0 = 1 | D_t^0 = 0, D_t^1 = 0)\,. \tag{7}$$

It is a direct treatment effect because we condition on D_t^a.

It is defined on the population of people who would have "survived" (i.e., not experienced the competing event) regardless of treatment.

Example: We can ask what the risk of death due to prostate cancer would have been at time t among those patients that would not have died due to any reasons other than prostate cancer, independent of whether they would have received estrogen therapy or placebo.

Counterfactual hazards under competing events

As in the case without competing events, we can look at hazards rather than risks. For example, the *hazard under elimination of competing events*, also called *marginal hazard*, is

$$P(Y_t^{a,\bar{d}_t=0} = 1 | Y_{t-1}^{a,\bar{d}_{t-1}=0} = 0).$$

An equivalent definition is, as in the case without competing events, related to the counterfactual survival time:

$$P(T^{a,\bar{d}_t=0} \in (t-1, t] \mid T^{a,\bar{d}_t=0} > t-1).$$

Similarly, we have the *hazard without elimination of competing events*, or *subdistribution hazard*, defined as

$$P(Y_t^a = 1 | Y_{t-1}^a = 0)\,.$$

Those individuals for which $Y_{t-1}^a = 0$ holds consist of both

1. individuals who did neither experience the event of interest nor the competing event and
2. individuals who who did not experience the event of interest, but the competing event.

Thus, an equivalent definition is:

$$P(T^a \in (t-1, t], \mathcal{D}^a = 1 | T^a > t-1 \text{ or } \{T^a \leq t-1 \text{ and } \mathcal{D}^a \neq 1\})$$

The *hazard conditioned on competing events*, or the *cause-specific hazard*, is the hazard of the event of interest among those who have not previously experienced the competing event:

$$P(Y_t^a = 1 | Y_{t-1}^a = D_t^a = 0)\,.$$

The above definitions give rise to the following possible contrasts:

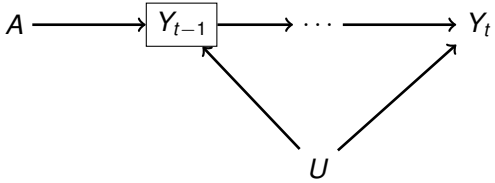
$$P(Y_t^{1,\bar{d}_t=0} = 1 | Y_{t-1}^{1,\bar{d}_{t-1}=0} = 0) - P(Y_t^{0,\bar{d}_t=0} = 1 | Y_{t-1}^{0,\bar{d}_{t-1}=0} = 0) \quad (8)$$

$$P(Y_t^1 = 1 | Y_{t-1}^1 = 0) - P(Y_t^0 = 1 | Y_{t-1}^0 = 0) \quad (9)$$

$$P(Y_t^1 = 1 | Y_{t-1}^1 = D_t^1 = 0) - P(Y_t^0 = 1 | Y_{t-1}^0 = D_t^0 = 0) \quad (10)$$

We could, of course, also look at ratios rather than differences.

The same considerations as in the case without competing events apply: a difference in hazards under $a = 1$ and $a = 0$ still does *not* necessarily imply that A has an effect on Y! This is because the hazards at t may differ simply because of different individuals who survive until $t - 1$ under $a = 1$ versus $a = 0$ due to treatment effects before time $t - 1$.



This is true for both the discrete-time and continuous-time hazard.

Censoring

We have so far considered (right-)censoring due to loss to follow-up, end of follow-up, transfer to different facilities etc.

Example: In the prostate cancer study, we considered censoring due to end of follow-up at 60 months. There is also censoring due to drop-out before 60 months.

All our estimands above are defined without referring to censoring so far.

As above, we can intervene on, condition on, or not specifically refer to censoring events and reflect this in our estimand definitions.

Loss to follow-up is typically considered to be a censoring event. In most cases, we would like to know the outcome that would have been observed without loss to follow-up, i.e. we'd like to intervene on it.

Is a competing event a censoring event?

Definition

A censoring event is any event occurring at t that ensures that the values of all future counterfactual outcomes under a are unknown even for an individual receiving the intervention a.

This means the chosen estimand determines whether the competing event is a censoring event or not.

Is a competing event a censoring event?

For the total effect (3), the subdistribution hazard contrast (9) and the cause-specific hazard contrasts (10), the competing event is *not* a censoring event because individuals with competing event at t, i.e. $(Y_t^a = 0, D_t^a = 1)$, have a known future counterfactual outcome: $Y_{t+1}^a = 0$.

Example: In the prostate cancer study, we may want to estimate the risk of death due to prostate cancer under estrogen therapy, without eliminating the competing event. For all patients, we know that if death due to any reason other than prostate cancer occurs at t, then death due to prostate cancer can not occur at $t + 1$. Thus, mortality due to other reasons is not a competing event under the above definition.

Is a competing event a censoring event?

For the direct effect (2) and the hazard contrast (8), for which we intervene on the competing event, the competing event *is* a censoring event. This is because individuals who experience a competing event at t have an unknown counterfactual future counterfactual outcome $Y_{t+1}^{a,\bar{d}_t=0}$.

Example: In the prostate cancer study, we may want to estimate the risk of death due to prostate cancer under estrogen therapy, if other deaths were not "allowed"[a]. For all patients, we know that death due to any reason other than prostate cancer can not occur at any t. The event indicator "death due to other reasons" at t does not tell us what $Y_{t+1}^{a,\bar{d}_t=0}$ would be. Thus, mortality due to other reasons is a competing event under the above definition.

[a] we discussed the complication of picking such an estimand already

Under which assumptions can we identify the estimands (2) - (10)?

Nothing new:

- consistency
- positivity
- conditional exchangeability

And we can check whether conditional exchangeability holds for a given causal model graphically, with a DAG.

We only need to consider the index t, the time ordering and reflect on which variables we intervene.

Conditional exchangeability

For 1 time point, we defined conditional exchangeability as:

$$Y^a \coprod A \mid L \quad \forall A = a, L = l \, .$$

We also said, that if L satisfies the back-door criterion, cond. exchangeability is achieved.

For multiple time points, we simply have statements in the spirit of

$$Y_t^{\text{interventions}} \coprod \text{all intervention variables} \mid \text{past}$$

Let's take the ordering (C_t, D_t, Y_t, L_t) at t, and stick to a single time point intervention variable $A = A_0$.

Everything we do also holds for $\bar{A}_t = (A_1, \ldots, A_t)$, but we keep it simple and stick to the prostate cancer example.

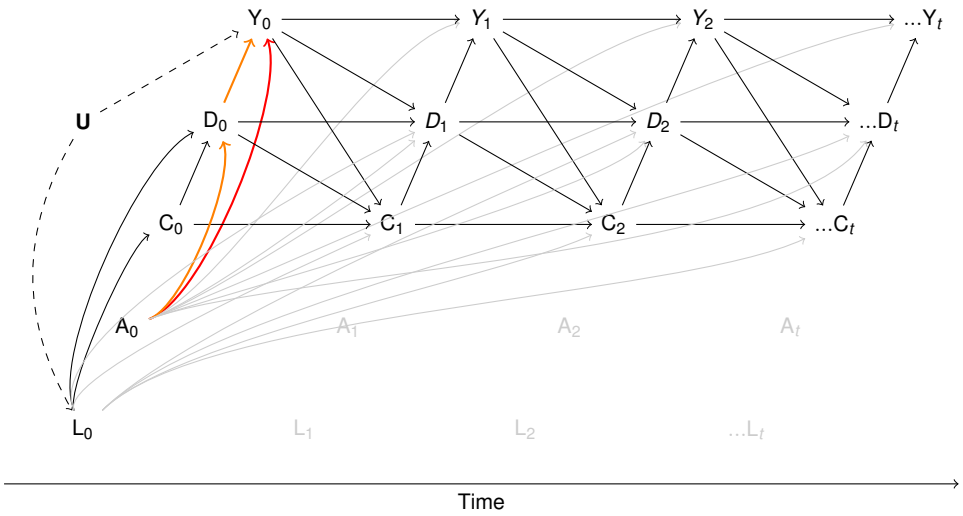
For the direct effect, defined in (2), the exchangeability statement is then

$$Y_t^{a,\bar{d}_t=\bar{c}_t=0} \amalg A \mid L_0 \tag{11}$$

$$Y_t^{a,\bar{d}_t=\bar{c}_t=0} \amalg C_t, D_t \mid \bar{L}_{t-1} = \bar{l}_{t-1}, \bar{Y}_{t-1} = \bar{C}_{t-1} = \bar{D}_{t-1} = 0, A = a \tag{12}$$

because we *intervene* on the censoring and competing event indicator!

Conditional exchangeability holds in this DAG, because no arrows from unmeasured variables into interventions (i.e. A, D, C)

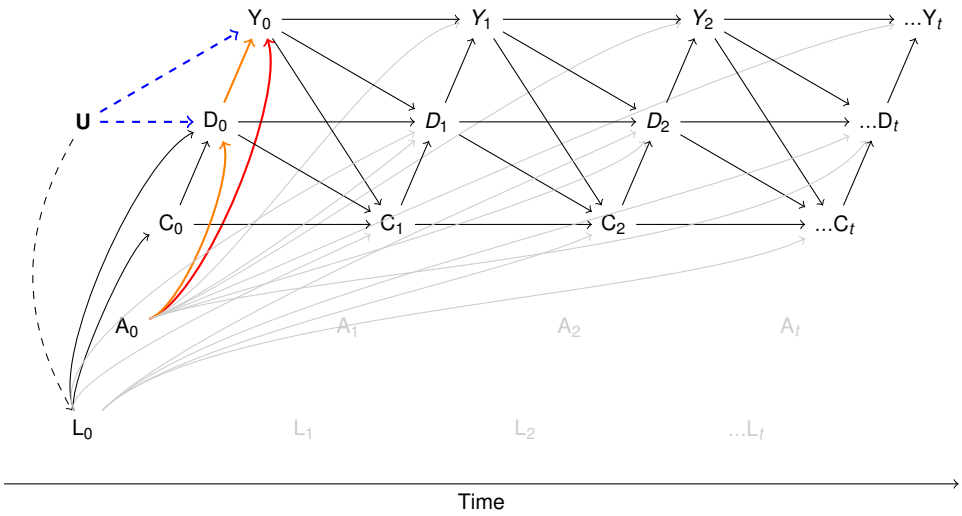


More on the DAG

- As is the prostate cancer example, we have only A_0 and L_0
- Some paths are in grey, for better readability; some are omitted (like $U \rightarrow Y_t$)
- We see one direct effect in red, and one indirect one in orange
- A_0 is randomized; no arrows into A_0; thus (11) holds
- Also (12) holds: no open back-door paths from D_t/C_t to Y_t
  $\rightarrow$ we have measured/adjusted for L_0 [2]

[2] there seems to be an open back-door path through C_0; but we intervene on C, it is thus a constant and in probability statements constants are always implicitly conditioned on.

Conditional exchangeability does not hold in this DAG, because of the arrow $U \rightarrow D_0$, and the associated back-door path



Simply ask yourself *why* people are being treated, censored or experience a competing event. If any of these reasons also cause, directly or indirectly, the outcome, then there is an open back-door path. It may not be needed to necessarily draw a DAG.

Example: In the prostate cancer study, the competing event is death due to reasons other than prostate cancer. There might be unmeasured factors, for instance comorbidities, that increase both the probability of death due to prostate cancer and due to other reasons. In this case, there would be unmeasured confounding, the direct effect can not be identified, and hence not consistently estimated. We would have to resort to a statistical interpretation as discussed in Chapter 8.

What about the risk without eliminating the competing event?

The same logic applies, but –as we discussed– the competing event is not a censoring event! We do not intervene on it.

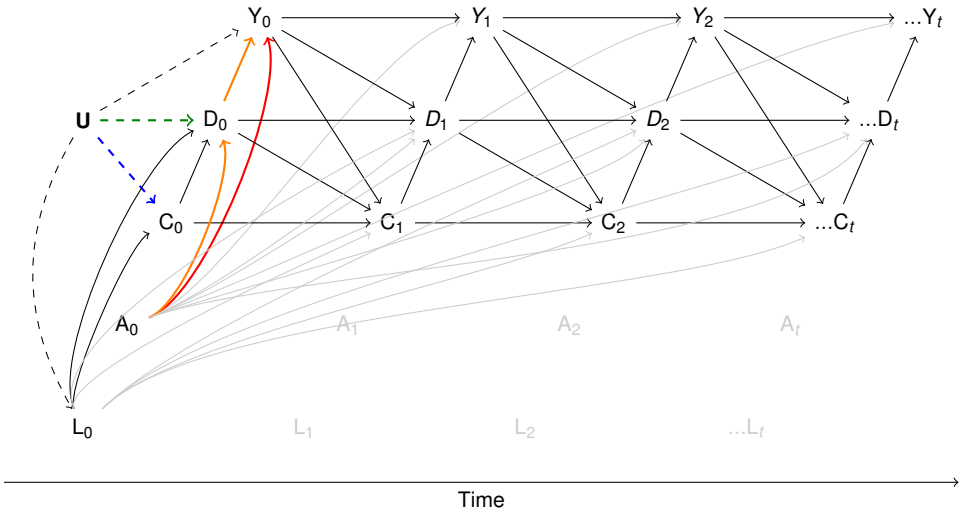
Thus, the identifying assumptions are weaker:

$$Y_t^{a,\bar{c}_t=0} \coprod A \mid L_0 \tag{13}$$

$$Y_t^{a,\bar{c}_t=0} \coprod C_t, \cancel{D_t} \mid \bar{L}_{t-1} = \bar{l}_{t-1}, \bar{D}_{t-1} = \bar{d}_{t-1}, \bar{Y}_{t-1} = \bar{C}_{t-1} = \cancel{\bar{D}_{t-1} = 0}, A = a \tag{14}$$

We are not requiring conditional exchangeability for D!

Conditional exchangeability holds with the green path; but not with the blue path



Example

Example: In the prostate cancer study, we can ask what the risk of death due prostate cancer would have been at time t, under either estrogen therapy and placebo, if we did not intervene on competing reasons of death. Identification of the corresponding estimand would work even if there are comorbidities that affect multiple reasons to die (prostate cancer and others). However, if there are reasons for drop-out (C_t) that also affect the event of interest, this would be a problem. While it is not impossible to think of such reasons, one could speculate that in the given study this is not a major concern.

For 1 time point, we defined positivity as

$$P(A = a|L = l) > 0 \quad \forall l \quad \text{with} \quad P(L = l) \neq 0\,. \tag{15}$$

and consistency as

$$\text{If } A = a, \text{ then } Y^a = Y \quad \text{for } \forall\, a\,;$$

For time-to-event data, we need similar statements but need not only refer to A, but also C_t and D_t if we intervene on them.

For the direct effect (2), we need both (15) to hold and

$$P(C_t = 0, D_t = 0 | \bar{L}_{t-1} = \bar{l}_{t-1}, C_{t-1} = D_{t-1} = Y_{t-1} = 0, A = a) > 0$$
$$\text{when} \quad f(a, \bar{l}_{t-1}, 0, 0, 0) \neq 0\,. \tag{16}$$

That is, we require that for any possible level of treatment and covariate history, among those that are uncensored, some individuals remain uncensored.

With additional parametric modeling assumptions, the assumption can sometimes be relaxed.

For the total effects, we only need to refer to C_t and not C_t and D_t.

Again, we are not going into more details.

For the direct effect (2), we require

$$\text{If } A = a, \quad \text{and} \quad \bar{C}_t = \bar{D}_t = 0 \text{ then } \bar{Y}_t^{a,\bar{c}=\bar{d}=0} = \bar{Y}_t \quad \text{and} \quad \bar{L}_t^{a,\bar{c}=\bar{d}=0} = \bar{L}_t \tag{17}$$

For the total effects, we only need to refer to C_t and not C_t and D_t.

What could be an issue with consistency in the context of the direct effect?

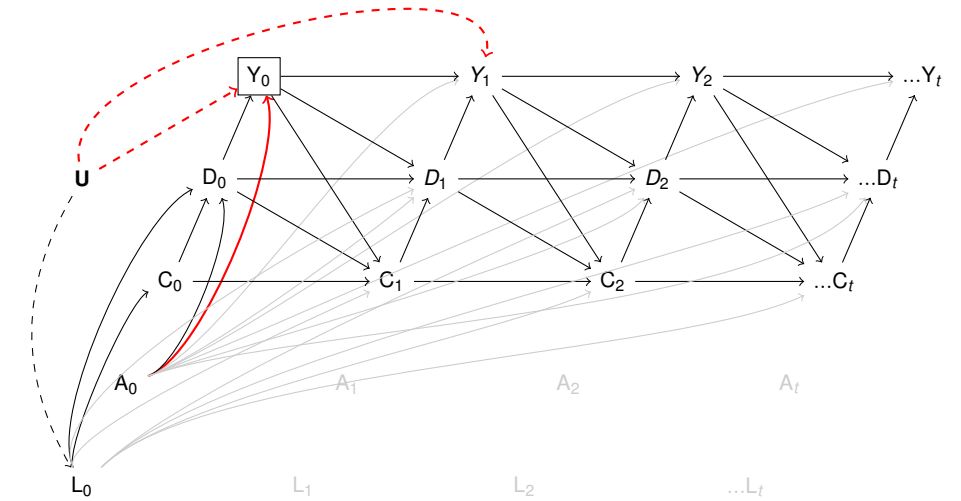
Again, we are not going into more details.

The same assumptions that allow identification of a direct effect on the event of interest (2) also give identification of a contrast in hazards under elimination of competing events (8).

The same assumptions that allow identification of the total effect on the event of interest (3) also give identification of the counterfactual contrast in subdistribution hazards (9).

The same assumptions that allow identification of the total effect on the event of interest (3), coupled with an additional set of assumptions that allow identification of the total effect on the competing event (4), allow identification of the counterfactual contrast in cause-specific hazards (10).

However, as we discussed already, contrasts of hazards typically do not have a causal interpretation in most realistic applications



When using the composite outcome to define effect measures, as in (5), there is no competing event anymore and identifying assumptions are weaker. As discussed, in the prostate example and many other examples, the combination of events into one outcome may not answer the question of interest; but there are a few cases where this estimand is a great alternative.

The principal stratum effect (7), i.e. SACE, can also be identified, but requires more complicated thoughts on “cross-worlds” and monotonicity. Its estimation is however not much more complicated than the other effect measures.

Theorem

Under conditional exchangeability, positivity and consistency –as defined in (11), (12), (15), (16), and (17)– it holds that the risk under elimination of competing and censoring events can be identified as

$$P(Y_t^{a,\bar{c}_t=\bar{d}_t=0}=1)=$$
$$\int_{\bar{\mathbf{l}}\in\bar{\mathbf{L}}_t}\left\{\begin{array}{l}\sum_{k=0}^{t}P(Y_k=1|Y_{k-1}=C_k=D_k=0,\bar{L}_{k-1}=\bar{l}_{k-1},A=a)\times\\ \prod_{j=0}^{k}P(Y_{j-1}=0|Y_{j-2}=C_{j-1}=D_{j-1}=0,\bar{L}_{j-2}=\bar{l}_{j-2},A=a)\times\\ f(l_j|\bar{l}_{j-1},Y_j=C_j=D_j=0,a)\end{array}\right\}d\bar{\mathbf{l}}\,. \tag{18}$$

More on the theorem

- The results looks complex, but in words it just says that we standardize (marginalize) with respect to the (post intervention) confounder distribution (as with a single time point), take the deterministic nature of C, D, Y into account and add up the risks over time. That's all.
- For the general case, we can replace A by $\bar{A}_t$.
- Proof: is long and thus omitted. It uses the same ingredients we used in the single time point case, i.e. laws of probability, conditional exchangeability, positivity, consistency
- For the total effects, we need to add one more factor, i.e. the conditional distribution of D_j, as we do not intervene on it.

- Estimation works by plugging into (18)
- We estimate the integral by simulation
- We do the same as in the single time point case:
  - go from left (past) to right (future) in the data
  - create counterfactual datasets where we set $A = a$ (e.g., 0/1)
  - simulate what would happen past the first intervention under $A = a$
- Only differences:
  - We intervene on > 1 variables
  - We have deterministic relationships between $C/D/Y$

1. Estimate outcome and confounder models at all t:
   a. Estimate $P(Y_t = 1 | Y_{k-1} = C_k = D_k = 0, \bar{L}_{k-1} = \bar{l}_{k-1}, A = a)$
   (among uncensored individuals at time t that did not have any event previously)
   b. Estimate the conditional distributions of the time-varying confounders L_t
   c. For total effects only: estimate the distributions of D_t conditional on the past
2. Create a counterfactual data set and set $A = 1$
   (keep data for pre-intervention variables, i.e. use empirical distributions)

3. Simulate data past (first) intervention:
   - generate stochastic draws of the conditional distributions from steps 1 forward in time, under the respective intervention $A = a$ by using the data from step 2
   - this produces a counterfactual data set
   - apply deterministic relationship: $Y_t = 1$ if $Y_{t-1} = 1$
     (for total effect: $D_t = 1$ if $D_{t-1} = 1$: $Y_t = 0$ if $D_{t-1} = 1$; $D_t = 0$ if $Y_{t-1} = 1$)
4. The mean of the estimated Y_t is then an estimate of $P(Y_t^{1,\bar{c}=\bar{d}=0} = 1)$
   (Total effect: the mean of the estimated D_t is an estimate of $P(D_t^{1,\bar{c}=0} = 1)$)
5. Repeat steps 2-4 for $A = 0$ and calculate the respective direct/total effects

Data with pre-intervention variables from Young et al. [1], in wide format:

	age	na	hx	hemoglobin	A	C.4	D.4	Y.4	C.8	D.8	Y.8	C.12	...
	...	..	..										...
247	71	0	1	14.09961	1	0	1	NA	NA	NA	NA	NA	...
248	73	1	1	13.59961	1	0	0	0	0	0	0	0	...
249	73	1	0	11.69922	0	0	0	0	0	0	0	0	...
250	68	1	0	13.39844	1	0	0	0	0	0	0	0	...
251	73	1	1	16.79688	1	0	0	0	0	1	NA	NA	...
252	82	1	1	12.39844	1	0	0	0	0	0	0	0	...

Estimation of direct effect with R (II)

```
# Step 1a: Fit models for Y_t; model specification based on variables used in
     paper, but with some model selection for bias-variance tradeoff
# NOTE: models are fit among those individuals who are uncensored and did
#       *not* experience any event yet

mY.4  <- gam(Y.4 ~ A+s(hemoglobin), data=prostate, family="binomial")
mY.8  <- gam(Y.8 ~ A+s(hemoglobin), data=prostate, family="binomial")
mY.12 <- gam(Y.12 ~ A+hx+s(hemoglobin), data=prostate, family="binomial")
...

# Step 1b: fit models for L_t: no time-varying confounders, so not needed
# Step 1c: fit models for D_t: not needed as we intervene on D_t
```


Estimation of direct effect with R (III)

```r
# Step 2:
# set A=1, leave L_0, and set everything post-intervention = NA
# C and D not needed as everyone should be uncensored in any case
sdata1 <- prostate
sdata1 <- subset(sdata1, select=-c(grep("C.",colnames(sdata1)),
                                   grep("D.",colnames(sdata1))))
sdata1[,grep("Y.",colnames(sdata1))] <- NA
sdata1$A <- 1
```

Data looks now as follows:

```
  age normal_activity hx hemoglobin A Y.4 Y.8 Y.12 ...
1  69               1  1   13.39844 1  NA  NA   NA ...
2  67               1  0   13.39844 1  NA  NA   NA ...
3  75               1  0   13.00000 1  NA  NA   NA ...
4  73               1  1   12.59961 1  NA  NA   NA ...
```


Estimation of direct effect with R (IV)

```r
# Step 3: simulate under A=1

sdata1$Y.4<-rbinom(n=252,1,prob=predict(mY.4,newdata=sdata1,type="response"))
sdata1$Y.8<-rbinom(n=252,1,prob=predict(mY.8,newdata=sdata1,type="response"))
...

# time-to-event data: deterministic rule: if Y_t-1=1, then Y_t=1
sdata1$Y.8[sdata1$Y.4==1]<-1
sdata1$Y.12[sdata1$Y.8==1]<-1
...
```

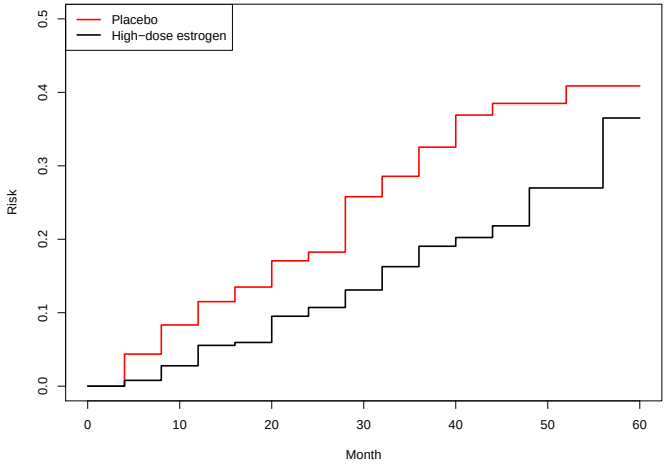
	age	na	hx	hemoglobin	A	Y.4	Y.8	Y.12	Y.16	Y.20	Y.24	Y.28	Y.32
1	69	1	1	13.39844	1	0	0	0	0	0	0	0	0
2	67	1	0	13.39844	1	0	0	0	0	0	0	0	0
3	75	1	0	13.00000	1	0	0	0	0	1	1	1	1
4	73	1	1	12.59961	1	0	0	0	0	0	0	0	0
5	74	1	1	13.59961	1	0	0	0	0	0	0	1	1
6	55	1	1	13.89844	1	0	0	0	0	0	0	0	0

Estimation of direct effect with R (V)

```r
# step 4: estimate E(Y_t^(a,c=0, d=0)) in counterfactual data
psi_1 <- apply(subset(sdata1, select=colnames(sdata1)[grep("Y.",colnames(
    sdata1))]) ,2,mean)

# step 5: repeat 2-4 for A=0 and draw curve of psi_1 and psi_0 over time
```


Results / Direct Effect



Estimation of total effect with R

```r
# Step 1a and b: as with direct effect
# Step 1c: Fit models for D_t
mD.4  <- gam(D.4 ~ A+hx+s(hemoglobin)+age, data=prostate, family="binomial")
mD.8  <- gam(D.8 ~ A+age, data=prostate, family="binomial")
...

# Step 2: intervene on C and A
#         as C=0, no specific action is needed
# set A=1, leave L_0, and set everything post-intervention = NA
sdata1 <- prostate
sdata1 <- subset(sdata1, select=-grep("C.",colnames(sdata1)) )
sdata1[,c(grep("Y.",colnames(sdata1)),grep("D.",colnames(sdata1)))] <- NA
sdata1$A <- 1
```


Estimation of total effect with R (II)

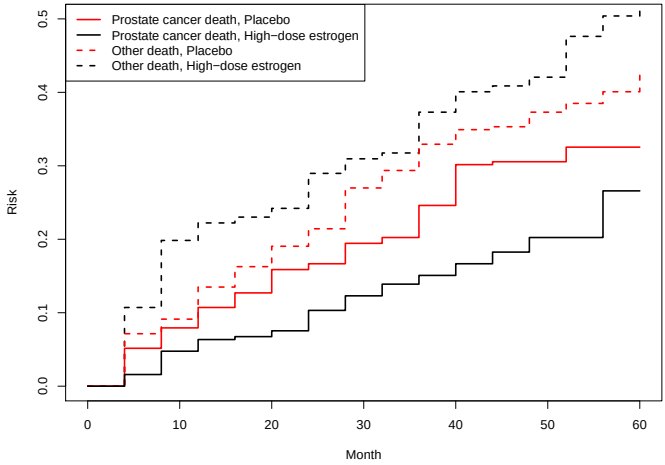
```
# Step 3
sdata1$D.4<-rbinom(n=252,1,prob=predict(mD.4,newdata=sdata1,type="response"))
sdata1$Y.4<-rbinom(n=252,1,prob=predict(mY.4,newdata=sdata1,type="response"))
sdata1$D.8<-rbinom(n=252,1,prob=predict(mD.8,newdata=sdata1,type="response"))
...

# if Y_t-1=1, then Y_t=1; and if D_t-1=1, then D_t=1
# also: if D_t=1, then Y_t=0 and if Y_t-1=1, then D_t=0
sdata1$Y.4[sdata1$D.4==1]<-0
sdata1$D.8[sdata1$D.4==1]<-1
sdata1$D.8[sdata1$Y.4==1]<-0
sdata1$Y.8[sdata1$Y.4==1]<-1
sdata1$Y.8[sdata1$D.8==1]<-0
...

# Step 4
psi_Y_1 <- apply(subset(sdata1, select=colnames(sdata1)[grep("Y.",colnames(
    sdata1))]) ,2,mean)
psi_D_1 <- apply(subset(sdata1, select=colnames(sdata1)[grep("D.",colnames(
    sdata1))]) ,2,mean)

# Step 5: repeat 2-4 for A=0
```


Results / Total Effect



- Estrogen therapy reduces mortality due to prostate cancer, but may increases mortality due to other reasons; if identification assumptions hold.
  (Bootstrap CI not shown)

- For the direct effect, assumptions are likely not met: we have only measured baseline confounders L_0, and no L_t; but there may be multiple unmeasured reasons of death due to other causes that also cause death due to prostate cancer, i.e. there is possibly some unmeasured confounding.

- It is possible that the beneficial effect of estrogen therapy on prostate cancer death is due to effects of therapy on other causes of death: when more people die from other causes, fewer can die from prostate cancer. How can we answer this? We need to look into the direct and indirect effects.
  $\rightarrow$ separable effects

Why is it plug-in estimation?

Example: Consider the prostate cancer study. For $t = 1$, setting $A = 1$, the ordering (C_t, D_t, Y_t) in each interval, treating L_0 as discrete and noting that $L_1 = \emptyset$, we can write (18) as

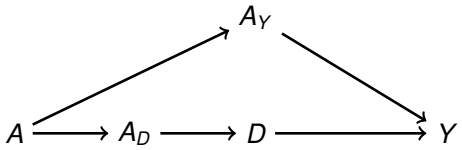
$$P(Y_1^{1,\bar{c}_1=\bar{d}_1=0} = 1) = \sum_{l_0} \begin{cases} P(Y_1 = 1 | Y_0 = C_0 = D_0 = C_1 = D_2 = 0, L_0 = l_0, A = 1) \times \\ P(Y_0 = 0 | C_0 = D_0 = 0, L_0 = l_0, A = 1) \times f(l_0) + \\ P(Y_1 = 0 | C_0 = D_0 = 0, L_0 = l_0, A = 1) \times f(l_0) \end{cases}$$

For estimation, we fit 2 outcome models. Then we use the empirical distribution for L_0, set $A = 1$ and predict the outcome under no censoring/competing events for the given L_0, A. Assume that $n = 100$ and 10% die at $t = 0$ and another 10% (i.e., $n = 10$) at $t = 1$. Then the estimated probability of death is simply the sum, i.e. 20%, and this is what we get when evaluating the counterfactual dataset. This is the same as calculating $(10/90) \cdot (1 - 0.1) + 10/100$. For those attending Lifetime Data Analysis, this looks familiar.

- We have learnt that competing events may often act es mediators.
- We also learnt how to define and estimate direct and total effects. Thus, we can now in principle understand the relevant pathways and if and how treatment works.
- The only problem we are left with is that in many situations interventions on the competing events, which we require to estimate the direct effect, do not make sense. For example, intervening on death is difficult to justify. In other situations, it may be o.k., like when “competing” events are defined as censoring-type of events (transfer to other programs, relocation) or very specific events (death by car accident).
- If there are such interpretational problems, we can address those by so-called separable effects – at least if it is possible to decompose the (biological) mechanisms through which treatment works.

Separable Effects

Suppose A acts on Y and D through different mechanisms and we describe those as A_D and A_Y:



Example: In the prostate cancer study, the estrogen diethylstilbestrol (DES, A) reduces prostate cancer mortality by suppressing testoserone production (A_Y); at the same time DES is believed to have negative effects, for example related to cardiac events due to complex biological mechanisms (A_D). We can think of interventions (e.g., other hormons) which reduce testoserone without having the same negative effects (on D), i.e. we intervene on A_Y only. Similarly, we can conceptualize interventions which target A_D only.

Definitions

We define

$$\begin{aligned} Y_t^{a_Y=a,a_D=a} &= Y_t^a, \\ D_t^{a_Y=a,a_D=a} &= D_t^a. \end{aligned}$$

The *separable direct effect* is

$$P(Y_t^{a_Y=1,a_d}) - P(Y_t^{a_Y=0,a_d}), \tag{19}$$

The *separable indirect effect* is

$$P(Y_t^{a_Y,a_d=1}) - P(Y_t^{a_Y,a_d=0}). \tag{20}$$

The separable direct and indirect effect add up to the total effect:

$$P(Y_t^{a=1}) - P(Y_t^{a=0}). \tag{21}$$

- For identification, we need essentially similar assumptions as before. For example, for the separable direct effect we still need conditional exchangeability as defined in (11) and (12), i.e. unmeasured common causes of both D_t and Y_t are still not allowed.

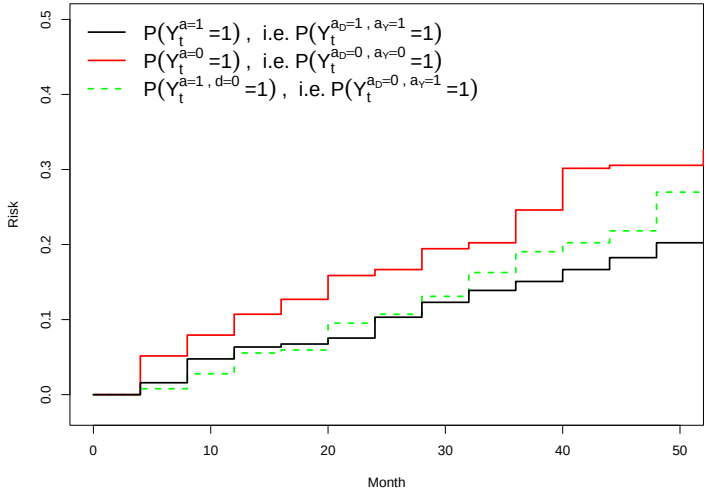
- Now, exchangeability need to be defined not only with respect to A, but both A_D and A_Y.

- Most importantly, the identification formula (18) is still valid as a basis for estimation of the *separable* direct effect. As before, for the total effect, we need to multiply the conditional density of D_t to the product term.
  Technically, the identification formula does not contain A, but A_Y (in the outcome components) and A_D (in the competing event component).

- As the identification formula is basically the same, the estimation procedures are the same.
- This means, there is nothing new we have to calculate: we already estimated the total and direct effect.
- The difference we get, is the difference in interpretation.
- We can use the results from before and plot them in 1 figure:

```
times <- seq(0,60,4)
plot(times,c(0,psi_1),type="s",lwd=2,lty=2, col="green")
lines(times,c(0,psi_Y_0),col="red",type="s",lwd=2)
lines(times,c(0,psi_Y_1),type="s",lwd=2)
```


Results



Interpretation

- DES reduces prostate cancer mortality

  $\rightarrow$ Total effect: black curve minus red curve.

- This is mostly[3] due to testoserone suppression (A_Y) because the indirect effect caused by mechanisms (A_D) for causes of death (D) is small (difference between black and green curve); thus, most of the effect may be attributed to the direct effect (difference between red and green).

  Separable indirect Effect: $P(Y^{a_D=1,a_Y=1} = 1) - P(Y^{a_D=0,a_Y=1} = 1)$

  Separable direct Effect: $P(Y^{a_D=0,a_Y=1} = 1) - P(Y^{a_D=0,a_Y=0} = 1)$

- The total effect of DES on prostate cancer mortality is not simply a consequence of a harmful effect on death from other causes!

- As discussed, the identification assumptions for the direct effect may not perfectly hold due to common comorbidities.

[3] for the first 35 weeks this is clear, afterwards not so much anymore. In Stensrud et al. [2] the results are clearer, likely owing to a more elaborate modeling approach

The hazard of hazard ratios

We already discussed caution around the interpretation of hazard ratios. So why bother again?

Because the standard survival regression models, used in epidemiology, give us hazard ratios!

The most commonly employed model is the proportional hazards model, typically attributed to Cox. Recall the continuous-time hazard from equation (1), which we denote as $\lambda(t)$. The Cox model, in its simplest form, is:

$$\lambda(t|L) \quad = \quad \lambda_0(t) \times \exp(\beta^T L)$$

where β is a vector of regression coefficients and $\lambda_0(t)$ is the baseline hazard.

The hazard of hazard ratios

The model parameters β are estimated with maximum partial likelihood estimation. Estimation details are discussed in Lifetime Data Analysis.

Suppose we only have 1 covariate L_1, which is binary. Then it is easy to see that

$$\frac{\lambda(t|1)}{\lambda(t|0)} = \exp(\beta_1),$$

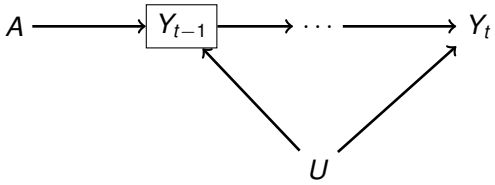
and hence the hazards are proportional. The hazard ratio does not vary over time, it is constant.

Example: In the prostate cancer study, fitting a Cox model with the variables used previously as covariates, yields $\beta_1 \approx -0.3$ (the coefficient associated with A) and $\exp(\beta_1) \approx 0.74$. This means, after adjusting for covariates, the mortality hazard is 0.73 times lower with treatment compared with no treatment. It is also possible to obtain an estimate of $\lambda_0(t)$; note however that the treatment effect does not vary over time (with this model specification).

The hazard of hazard ratios

1. Many studies report only a single hazard ratio, obtained from a Cox, or similar, model. This may be uninformative because the hazard ratio may be time-varying; in fact, it is often argued that most epidemiological studies have treatment effects that change over time [3]. A single HR is a weighted average of the time-varying hazard ratios. The cumulative benefit at a particular time point can only be conveyed by a comparison of *risks* in each group.

2. This problem can be fixed easily by reporting period-specific hazard ratios, using refined Cox model specifications or the estimation methods we discussed earlier, but then we likely face the built-in selection bias, as discussed:



The hazard of hazard ratios in competing risks settings

In competing risks settings, it is often recommended to use regression models, that

1. either model the *conditional* cause-specific hazards $\lambda_k(t|L)$ for event $D = k$, where

$$\lambda_k(t) \quad = \quad \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t < T \leq t + \Delta t, D = k \mid T > t)$$

2. or the *conditional* subdistribution hazard $\lambda_k^*(t|L)$ with

$$\lambda_k^*(t) \quad = \quad \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t < T \leq t + \Delta t, D = k \mid T > t \text{ or } (T \leq t \text{ and } D \neq k)) .$$

We can estimate (1) by using the Cox model and standard software, when censoring observations if they experience a competing event. We can estimate (2) with specific packages, e.g. `cmprsk` in *R*.

Reflections

- When we measure the right variables, and achieve conditional exchangeability, we can in principle get causal estimands through these models. Those are conditional on L, rather than marginal, but let's ignore this for now.
- But the logic from before applies again: we will face the built-in selection bias whenever there are unmeasured common causes of Y_{t-1} and Y_t (i.e. almost always in epidemiology).
- Adjusted risk/survival curves, as we constructed, avoids this problem and can be easily understood.
- Under purely descriptive aims, we can simply plot cumulative incidences rather than fitting models. Then we can tell what proportion of individuals experienced event $D = k$ *before* any other event occurred. Estimation can be facilitated by the g-formula, as we did, or typically differently (somewhat simpler).
- Of course, there are situations where the survival bias is small, the subdistribution hazard is a good summary measure over time and the model hence useful.

Summary

- In time-to-event analyses, competing events may prohibit individuals to experience the event of interest
- A fundamental insight is that the competing event may be a mediator on the path from A to Y.
- The simplest way to fix this, is to combine Y and D into one outcome. There is absolutely nothing wrong in doing this, but in many cases this may not answer the question of interest (e.g. prostate cancer study). An example where this works would be the combined outcome (AIDS [=clincally severe event], death), see target trial from chapter 6.

Summary

- Historically, classic estimands in competing risk settings relate to hazards. We have learnt that hazards as estimands come with a built-in selection bias in most realistic cases, which is not ideal when we want to learn about (time-varying) treatment effects.
- Estimands that are based on risks avoid this problem.
- We saw that we could condition, intervene or not intervene on the competing event.
- Conditioning on competing events, can easily yield collider bias; whereas intervening or not intervening are viable options.
- That is, we concluded that the risk under elimination of competing events as well as the cause-specific cumulative incidences are good target estimands.

Summary

- We need less assumptions to identify the cumulative incidences (risk without intervening); we can –in particular– have unmeasured common causes of the outcome and the competing event. Thus this would be the most preferable estimand to pick.

- To give the most nuanced interpretations, it would be good to additionally estimate the direct (and indirect) effect; if this can be somehow defended. Then we know for which reasons we observe treatment effects. A conceptualization through seperable effects can help here to disentangle the mechanisms at work.

- Estimation through the g-formula always works. We discussed already some disadvantages (bootstrapping, model specifications). However, there is no time to discuss alternatives.

Bibliography

[1] Jessica G. Young, Mats J. Stensrud, Eric J. Tchetgen Tchetgen, and Miguel A. Hernán.
A causal framework for classical statistical estimands in failure-time settings with competing events.
Statistics in Medicine, 39(8):1199–1236, 2020.

[2] M. J. Stensrud, J. G. Young, V. Didelez, J. M. Robins, and M. A. Hernan.
Separable effects for causal inference in the presence of competing events.
Journal of the American Statistical Association, 117(537):175–183, 2022.

[3] Mats J. Stensrud and Miguel A. Hernán.
Why Test for Proportional Hazards?
JAMA, 323(14):1401–1402, 2020.

[4] M. Hernan and J. Robins.
Causal inference.
Chapman & Hall/CRC, Boca Raton, 2022.

[5] M. A. Hernan.
The hazards of hazard ratios.
Epidemiology, 21(1):13–15, 2010.