Causal Inference with competing events

Mats Stensrud, Jessica Young & L. Paloma Rojas-Saunero

École Polytechnique Fédérale de Lausanne (EPFL)

Department of Mathematics

Analyses are **practically useful** if they correspond to **practically relevant questions**.[1][2][3]

[1] Jessica G. Young et al. "A causal framework for classical statistical estimands in failure-time settings with competing events". In: *Statistics in Medicine* 39.8 (2020), pp. 1199–1236. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8471.

[2] Mats J Stensrud and Oliver Dukes. "Translating questions to estimands in randomized clinical trials with intercurrent events". In: *Statistics in Medicine* (2022).

[3] Aaron L Sarvet and Mats J Stensrud. "Without Commitment to an Ontology, There Could Be No Causal Inference". In: *Epidemiology* 33.3 (2022), pp. 372–378.

Analyses are **practically useful** if they correspond to **practically relevant questions**.[1][2][3]

$\Longrightarrow$ We should be precise about – and justify – the estimand, that is, the formalization of our research question.

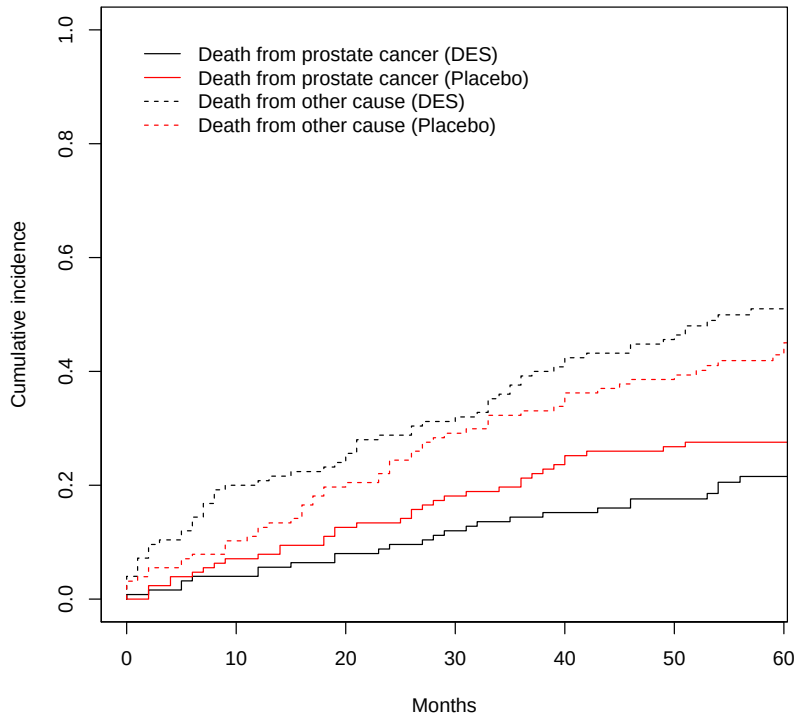
[1] Young et al., "A causal framework for classical statistical estimands in failure-time settings with competing events".

[2] Stensrud and Dukes, "Translating questions to estimands in randomized clinical trials with intercurrent events".

[3] Sarvet and Stensrud, "Without Commitment to an Ontology, There Could Be No Causal Inference".

A drug to reduce prostate cancer mortality

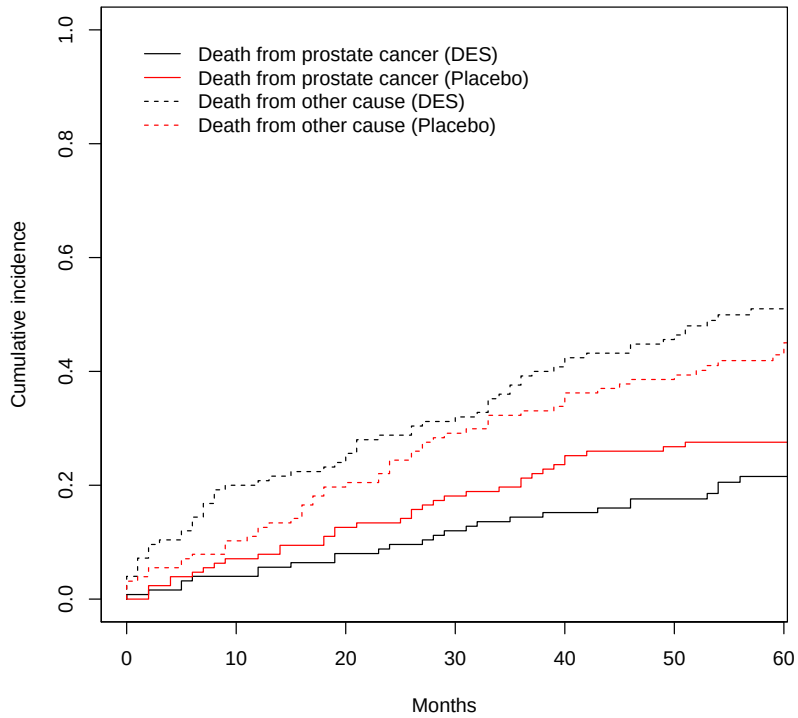
`Byar and Green (1980)`



- Diethylstilbestrol (DES, *black*) vs Placebo (*red*).

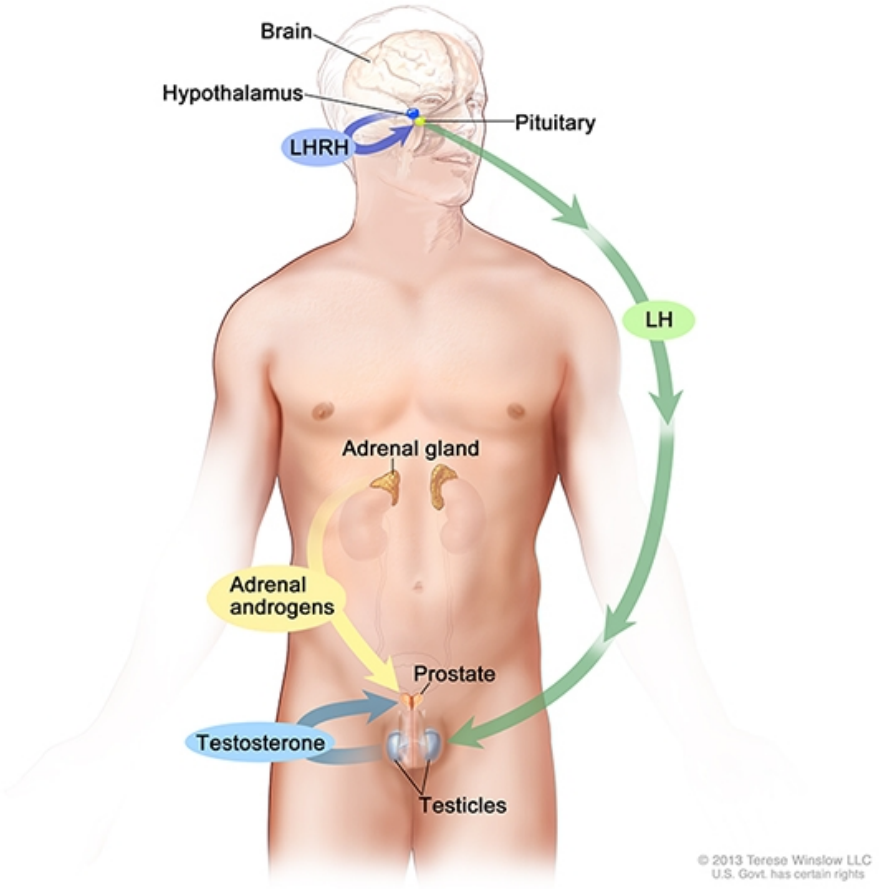
A drug to reduce prostate cancer mortality

`Byar and Green (1980)`



- Diethylstilbestrol (DES, *black*) vs Placebo (*red*).
- When more people die from other causes, fewer people can die from prostate cancer.

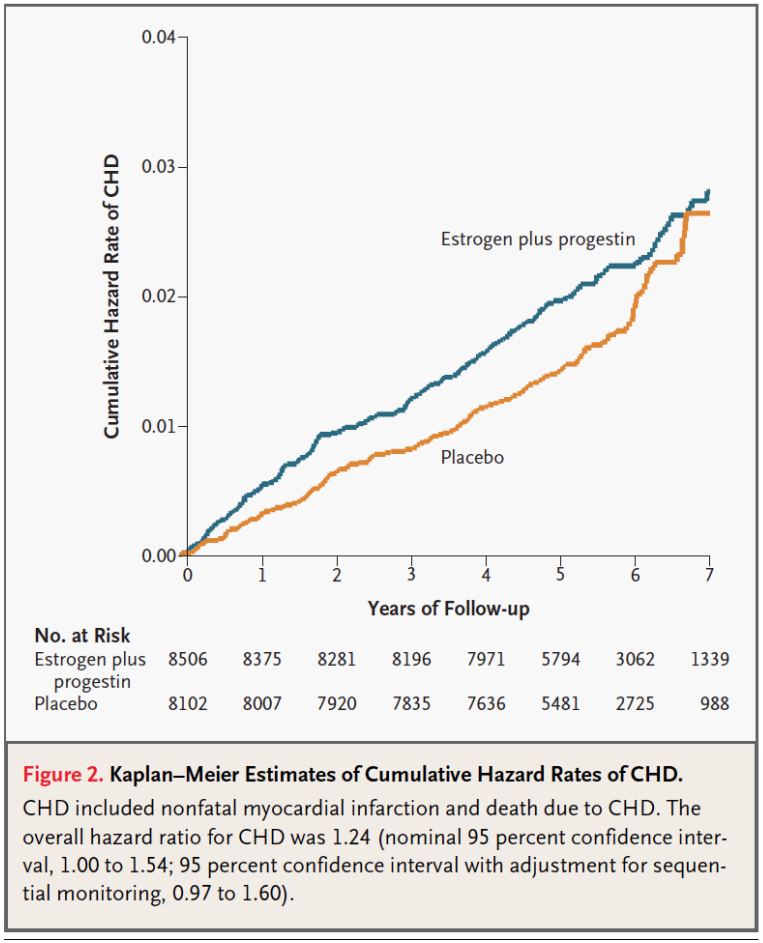
Prostate example: How does DES work?



`https://www.cancer.gov/types/prostate`

Stops testosterone production but increases cardiovascular risk?

Postmenopausal Estrogen (NEJM)



No. at Risk								
Estrogen plus progestin	8506	8375	8281	8196	7971	5794	3062	1339
Placebo	8102	8007	7920	7835	7636	5481	2725	988

Figure 2. Kaplan–Meier Estimates of Cumulative Hazard Rates of CHD.
CHD included nonfatal myocardial infarction and death due to CHD. The overall hazard ratio for CHD was 1.24 (nominal 95 percent confidence interval, 1.00 to 1.54; 95 percent confidence interval with adjustment for sequential monitoring, 0.97 to 1.60).

Manson et al, NEJM 2003

This is easy. We have a strategy:

- Censor those who died from the competing event.

This is easy. We have a strategy:

- Censor those who died from the competing event.
- Fit a Cox model.

This is easy. We have a strategy:

- Censor those who died from the competing event.
- Fit a Cox model.
- Report a hazard ratio.

Not so fast...

- Censor those who died of the competing event.

  Which scientific question are we targeting?
  What's our estimand?

  Robins, Mathematical Modelling, 1986.
  Young, Stensrud, Tchetgen Tchetgen & Hernán, Stat Med, 2020.

- Fit a Cox model.

  Are the hazards ever proportional?
  If they're not proportional, what's the consequence for inference?

  Stensrud & Hernán, JAMA, 2020.

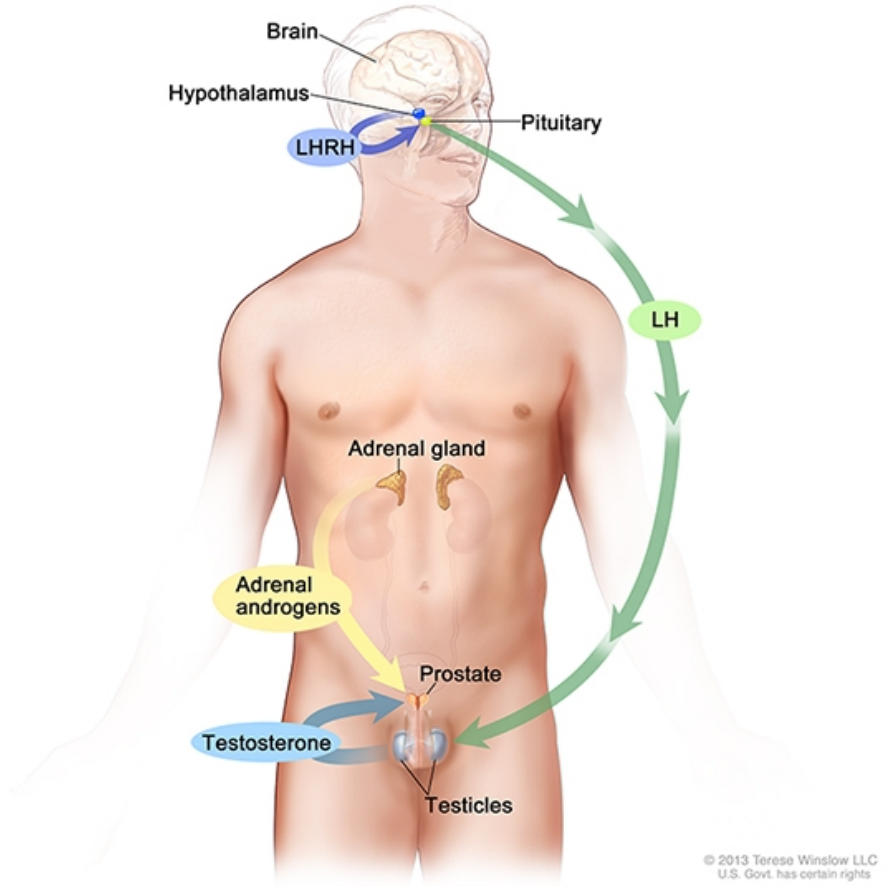
- Report a hazard ratio.

  How do we interpret the hazard ratio anyway?

Observed Data Structure (reminder)

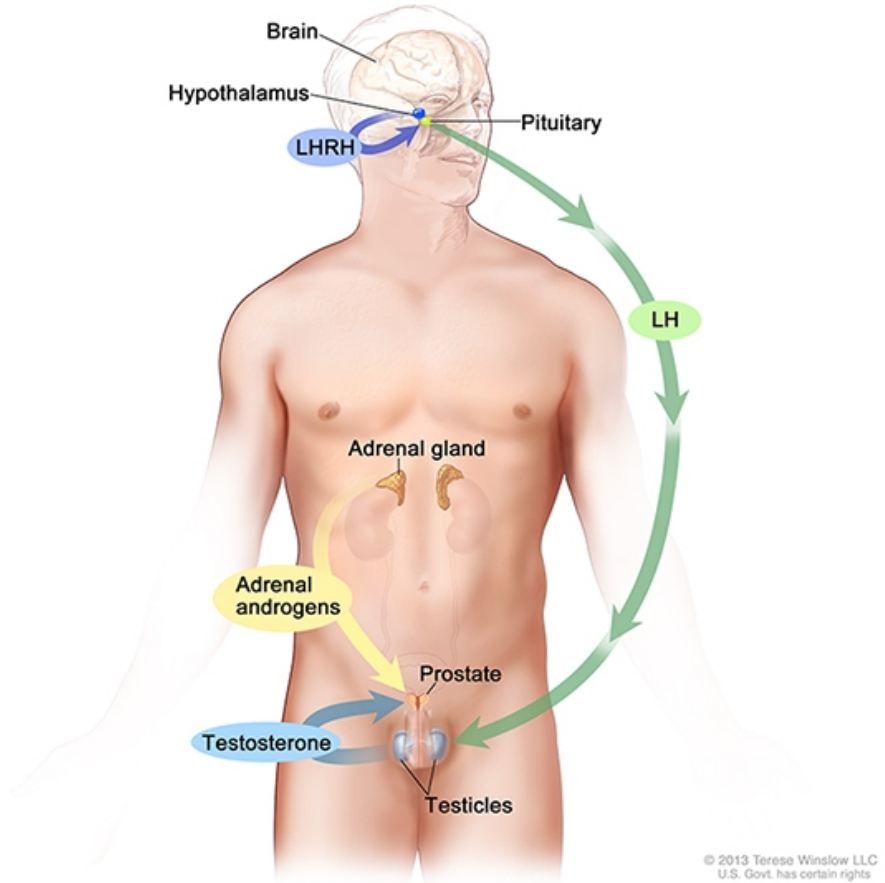
- Binary treatment $A \in \{0, 1\}$.
- Discrete time intervals $k = 0, 1, 2..., K + 1$.
- Event of interest Y_k (death due to prostate cancer).
- Competing event D_k (death from other causes).
- Baseline covariates Z.
- Censoring C_k.
- Time-dependent covariates Z_k.
- Continuous time.

Prostate example: How does DES work?



https://www.cancer.gov/types/prostate

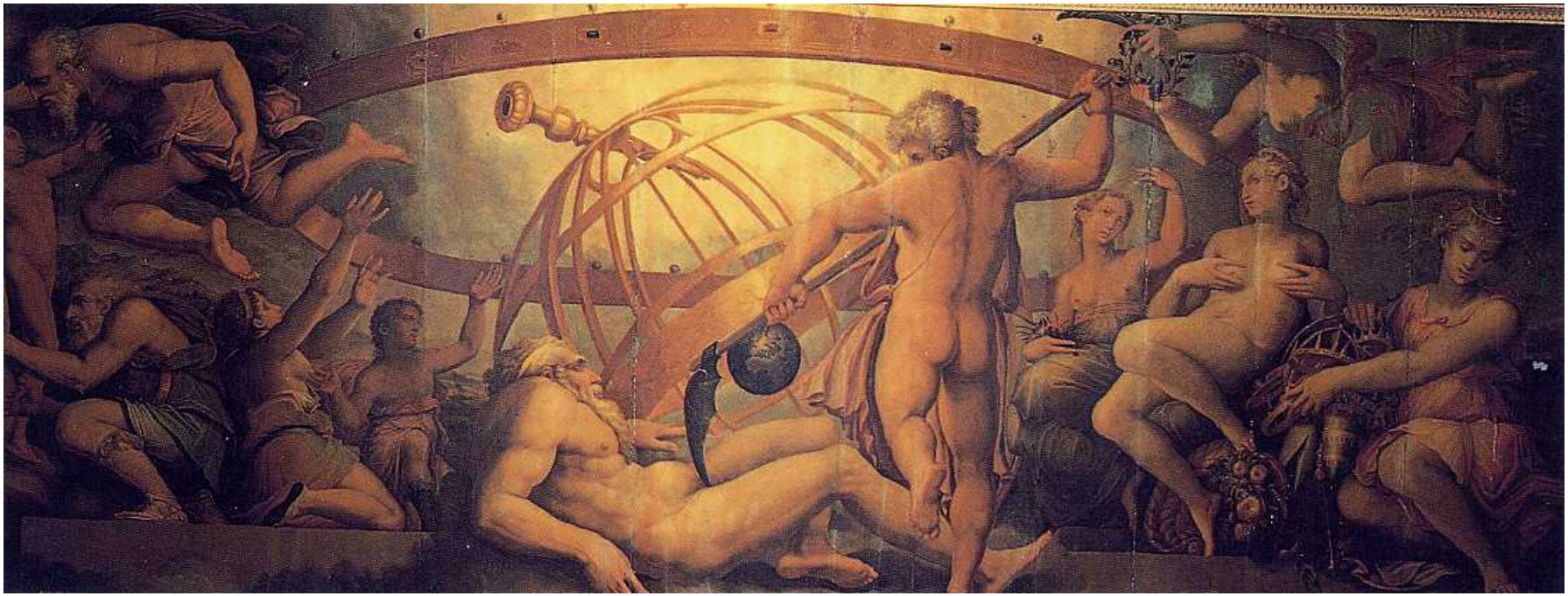
Prostate example: How does DES work?



`https://www.cancer.gov/types/prostate`

Stops testosterone production but increases cardiovascular risk?

Alternatives...



The Mutilation of Uranus by Saturn...

Modified treatment assumption

Consider a new study where, instead of assigning A, we assign combinations of **two new binary treatments A_Y and A_D**.

Modified treatment assumption

Consider a new study where, instead of assigning A, we assign combinations of **two new binary treatments A_Y and A_D**.

- Modified treatment assumption: Jointly assigning A_Y and A_D to the same value a leads to exactly the same values of Y_k and $D_k, k \in \{0, \ldots, K\}$ had we assigned A to a.

Modified treatment assumption

Consider a new study where, instead of assigning A, we assign combinations of **two new binary treatments A_Y and A_D**.

- Modified treatment assumption: Jointly assigning A_Y and A_D to the same value a leads to exactly the same values of Y_k and $D_k, k \in \{0, \ldots, K\}$ had we assigned A to a.
- Need this assumption to:
  - identify the separable effects, and
  - interpret them as capturing direct, indirect or path-specific effects of A.

Modified treatment assumption

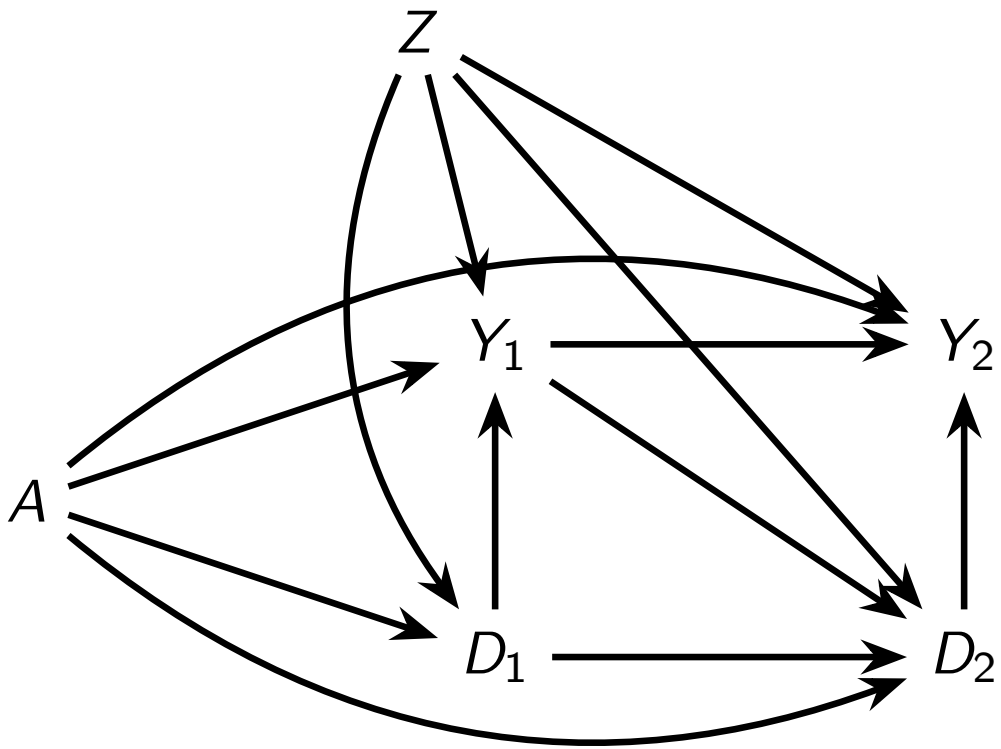
Consider a new study where, instead of assigning A, we assign combinations of **two new binary treatments A_Y and A_D**.

- Modified treatment assumption: Jointly assigning A_Y and A_D to the same value a leads to exactly the same values of Y_k and $D_k, k \in \{0, \ldots, K\}$ had we assigned A to a.
- Need this assumption to:
  - identify the separable effects, and
  - interpret them as capturing direct, indirect or path-specific effects of A.

This assumption could in principle be falsified in a 6-arm trial, which includes all joint combinations of A_Y and A_D plus each level of A.

- Example where this will hold is when A_Y and A_D are a physical decomposition of A.

Illustrative Directed Acyclic Graph (DAG)



Causal DAG with two time intervals.

Observed Data Structure

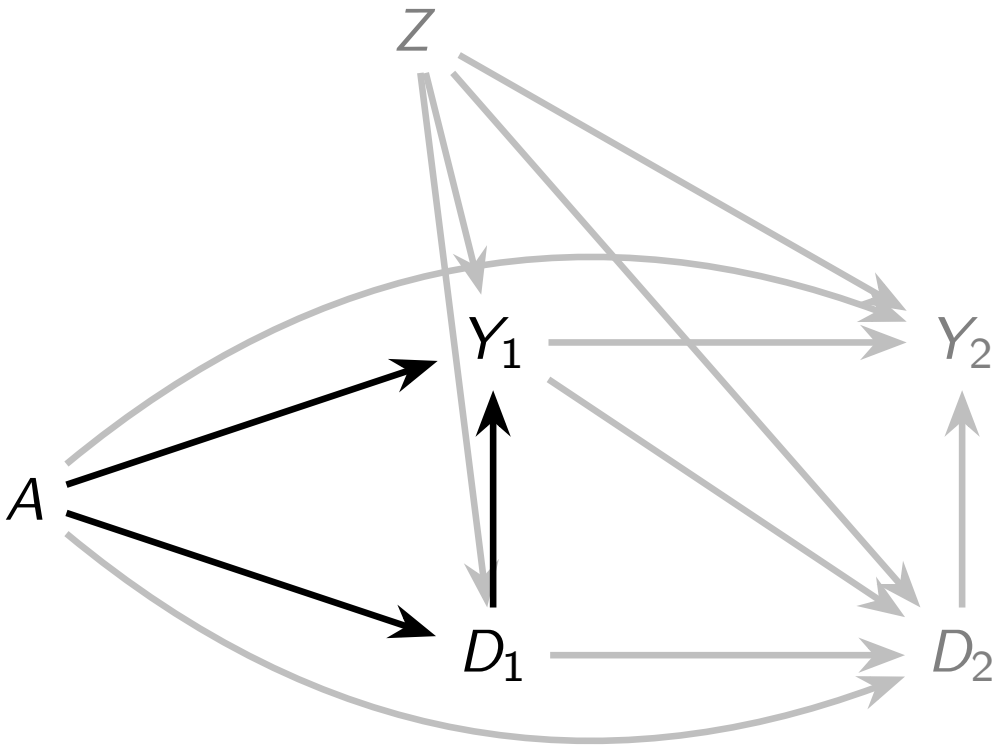
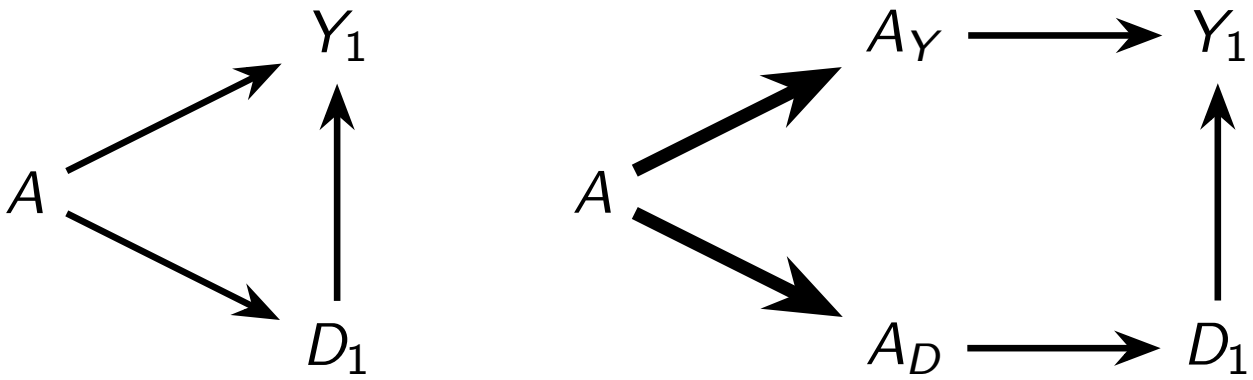


Figure: Causal Directed Acyclic Graph

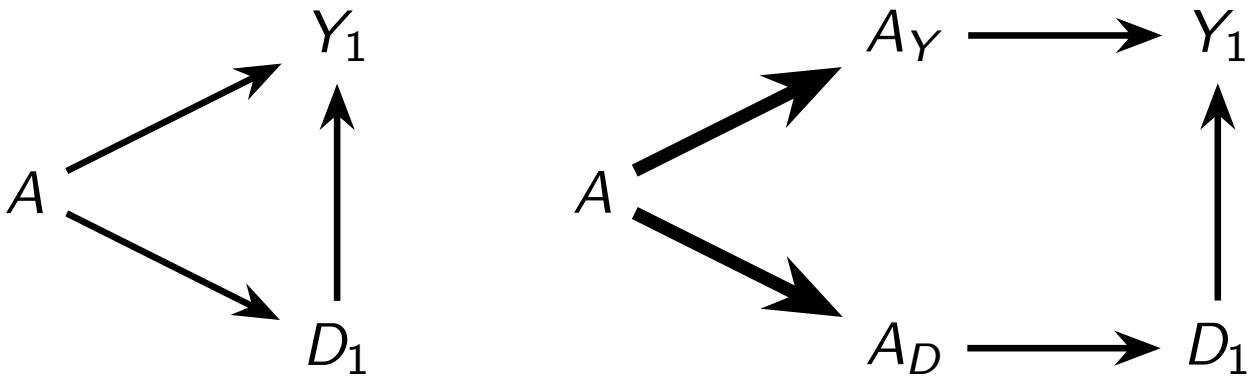
Decomposition of A in a simple setting



(a) (b)

Figure: Bold arrows denote deterministic relationships.
`Inspired by Robins and Richardson (2010)`

Decomposition of A in a simple setting

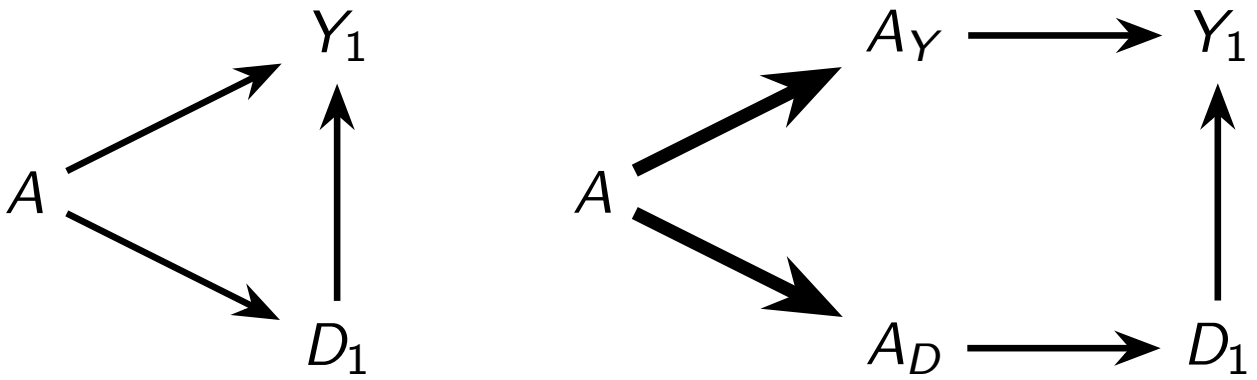


(a) (b)

Figure: Bold arrows denote deterministic relationships.
`Inspired by Robins and Richardson (2010)`

- A can be decomposed into $A_Y \in \{0, 1\}$ and $A_D \in \{0, 1\}$ such that, $A \equiv A_D \equiv A_Y$.

Decomposition of A in a simple setting



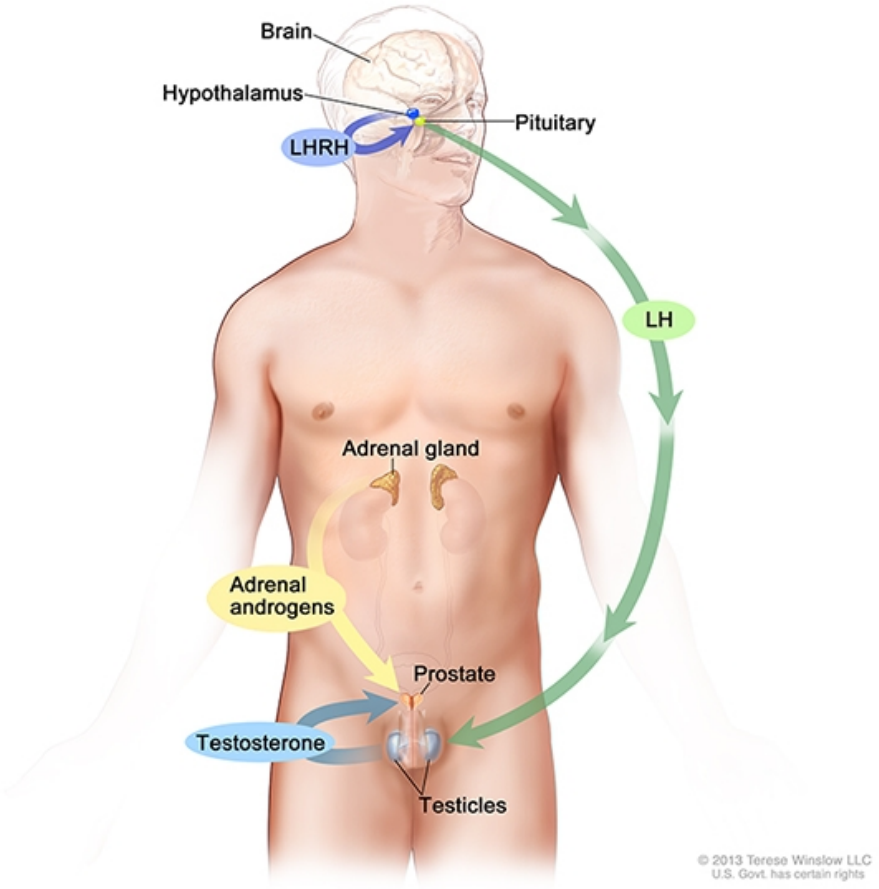
(a) (b)

Figure: Bold arrows denote deterministic relationships.
`Inspired by Robins and Richardson (2010)`

- A can be decomposed into $A_Y \in \{0, 1\}$ and $A_D \in \{0, 1\}$ such that, $A \equiv A_D \equiv A_Y$.
- Assigning $A = a$ results in the same outcomes as an intervention that assigns $(A_Y = a, A_D = a)$,

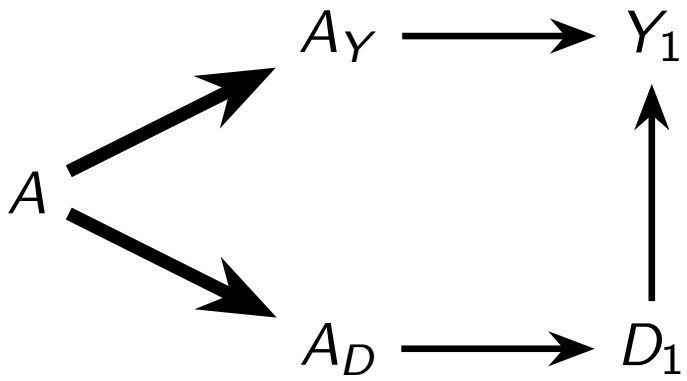
$$Y_{k+1}^{a_Y = a_D = a} = Y_{k+1}^{a}.$$

Prostate example: How does DES work?



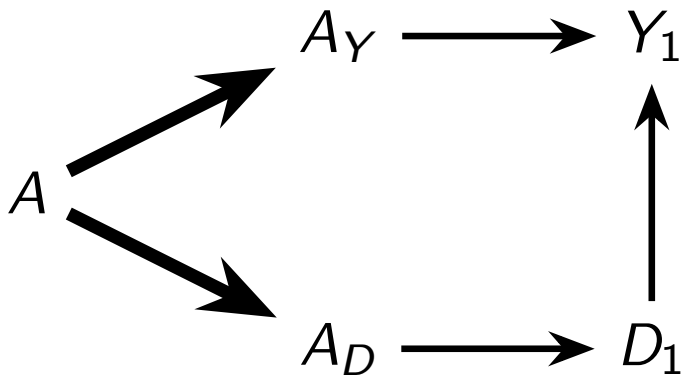
https://www.cancer.gov/types/prostate

Isolation assumptions



Informally, no direct effect of A_D on Y_k and of A_Y on D_k.

Isolation assumptions

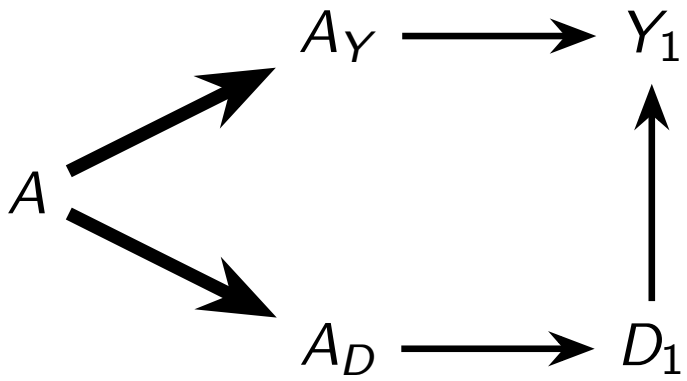


Informally, no direct effect of A_D on Y_k and of A_Y on D_k.
More formally, full isolation is satisfied if:

- the only causal paths from A_Y to D_k $k = 0, ..., K$ are directed paths intersected by Y_{j-1}, $j = 0, ..., k$,

Stensrud et al, 2020a, JASA Theory & Methods

Isolation assumptions

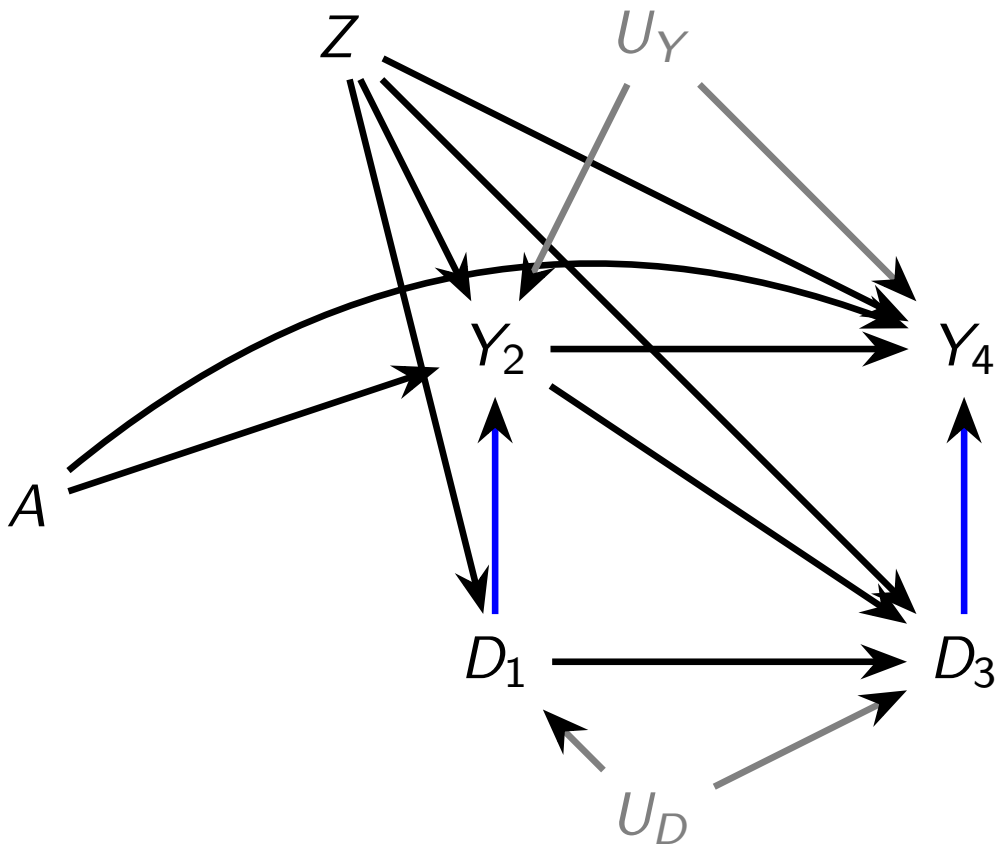


Informally, no direct effect of A_D on Y_k and of A_Y on D_k.
More formally, full isolation is satisfied if:

- the only causal paths from A_Y to D_k $k = 0, ..., K$ are directed paths intersected by Y_{j-1}, $j = 0, ..., k$,
- the only causal paths from A_D to Y_k, $k = 0, ..., K$ are directed paths intersected by D_j, $j = 0, ..., k$.

Stensrud et al, 2020a, JASA Theory & Methods

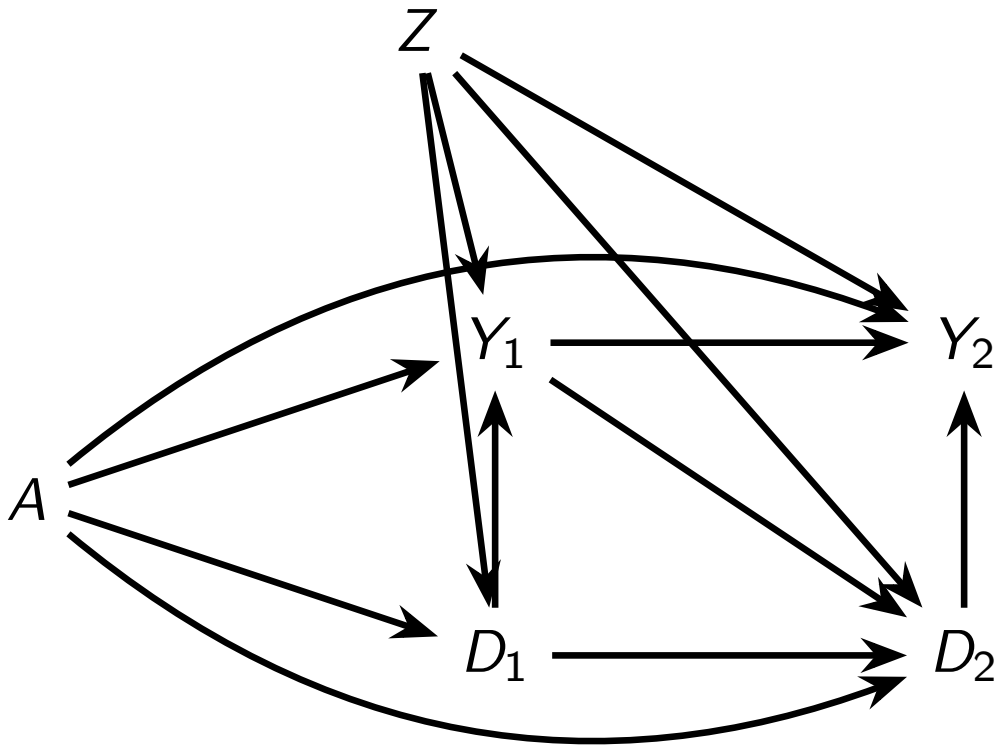
Example: total effect is a "direct effect"



A is PrEP for HIV, Y is HIV incidence, D is death.

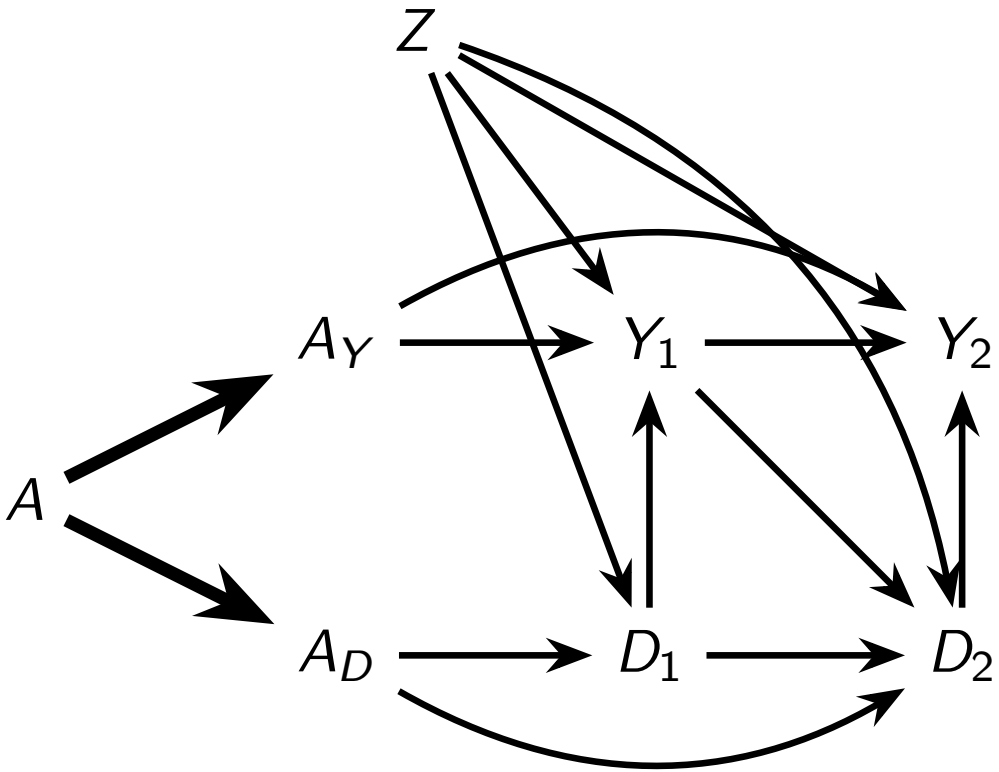
Reasonable to assume that treatment affects death only via its protection against HIV incidence. Here A_D is empty.

DAG before decomposition of A



Causal graph that describes two time intervals.

Extended DAG



Extended causal graph.

Defining separable effects

Defining separable effects

- Separable direct effects:

$$\Pr(Y_k^{a_Y=1,a_D} = 1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D} = 1), \quad a_D \in \{0, 1\}$$

Defining separable effects

- Separable direct effects:

$$\Pr(Y_k^{a_Y=1,a_D} = 1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D} = 1), \quad a_D \in \{0, 1\}$$

Effect of A on Y_k not through $\bar{D}_k$.

Defining separable effects

- Separable direct effects:

$$\Pr(Y_k^{a_Y=1,a_D} = 1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D} = 1), \quad a_D \in \{0,1\}$$

Effect of A on Y_k not through $\bar{D}_k$.

- Separable indirect effects:

$$\Pr(Y_k^{a_Y,a_D=1} = 1) \text{ vs. } \Pr(Y_k^{a_Y,a_D=0} = 1), \quad a_Y \in \{0,1\}$$

Defining separable effects

- Separable direct effects:

$$\Pr(Y_k^{a_Y=1,a_D} = 1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D} = 1), \quad a_D \in \{0,1\}$$

Effect of A on Y_k not through $\bar{D}_k$.

- Separable indirect effects:

$$\Pr(Y_k^{a_Y,a_D=1} = 1) \text{ vs. } \Pr(Y_k^{a_Y,a_D=0} = 1), \quad a_Y \in \{0,1\}$$

Effect of A on Y_k only through $\bar{D}_k$.

Separable effects sum to the total effect

$$
\begin{aligned}
&\text{Separable direct effect} + \text{Separable indirect effect} \\
=&[\Pr(Y_k^{a_Y=1,a_D=1}=1) - \Pr(Y_k^{a_Y=0,a_D=1}=1)] \\
&+ [\Pr(Y_k^{a_Y=0,a_D=1}=1) - \Pr(Y_k^{a_Y=0,a_D=0}=1)] \\
=&\Pr(Y_k^{a=1}=1) - \Pr(Y_k^{a=0}=1).
\end{aligned}
$$

The Systolic Blood Pressure Intervention Trial (NEJM, 2015)

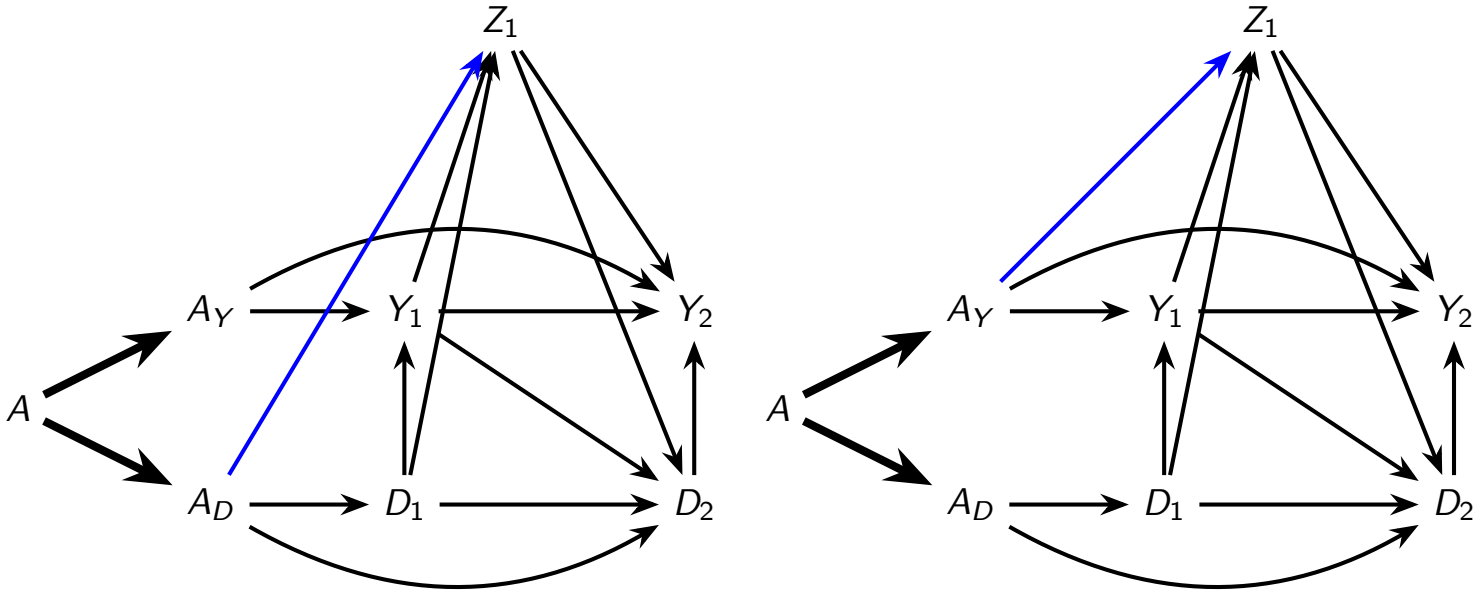
Table 3. Serious Adverse Events, Conditions of Interest, and Monitored Clinical Events.

Variable	Intensive Treatment (N=4678)	Standard Treatment (N=4683)	Hazard Ratio	P Value
	no. of patients (%)			
Serious adverse event*	1793 (38.3)	1736 (37.1)	1.04	0.25
Conditions of interest				
Serious adverse event only				
Hypotension	110 (2.4)	66 (1.4)	1.67	0.001
Syncope	107 (2.3)	80 (1.7)	1.33	0.05
Bradycardia	87 (1.9)	73 (1.6)	1.19	0.28
Electrolyte abnormality	144 (3.1)	107 (2.3)	1.35	0.02
Injurious fall†	105 (2.2)	110 (2.3)	0.95	0.71
Acute kidney injury or acute renal failure‡	193 (4.1)	117 (2.5)	1.66	<0.001
Emergency department visit or serious adverse event				
Hypotension	158 (3.4)	93 (2.0)	1.70	<0.001
Syncope	163 (3.5)	113 (2.4)	1.44	0.003
Bradycardia	104 (2.2)	83 (1.8)	1.25	0.13
Electrolyte abnormality	177 (3.8)	129 (2.8)	1.38	0.006
Injurious fall†	334 (7.1)	332 (7.1)	1.00	0.97
Acute kidney injury or acute renal failure‡	204 (4.4)	120 (2.6)	1.71	<0.001
Monitored clinical events				
Adverse laboratory measure§				
Serum sodium <130 mmol/liter	180 (3.8)	100 (2.1)	1.76	<0.001
Serum sodium >150 mmol/liter	6 (0.1)	0		0.02
Serum potassium <3.0 mmol/liter	114 (2.4)	74 (1.6)	1.50	0.006
Serum potassium >5.5 mmol/liter	176 (3.8)	171 (3.7)	1.00	0.97
Orthostatic hypotension¶				
Alone	777 (16.6)	857 (18.3)	0.88	0.01
With dizziness	62 (1.3)	71 (1.5)	0.85	0.35

Observed Data Structure 2

- Binary treatment $A \in \{0, 1\}$ (blood pressure therapy).
- Discrete time intervals $k = 0, 1, 2..., K$.
- Event of interest Y_k (acute kidney injury).
- Competing event D_k (death).
- Baseline covariates Z_0.
- Time-varying covariates Z_k (blood pressure).

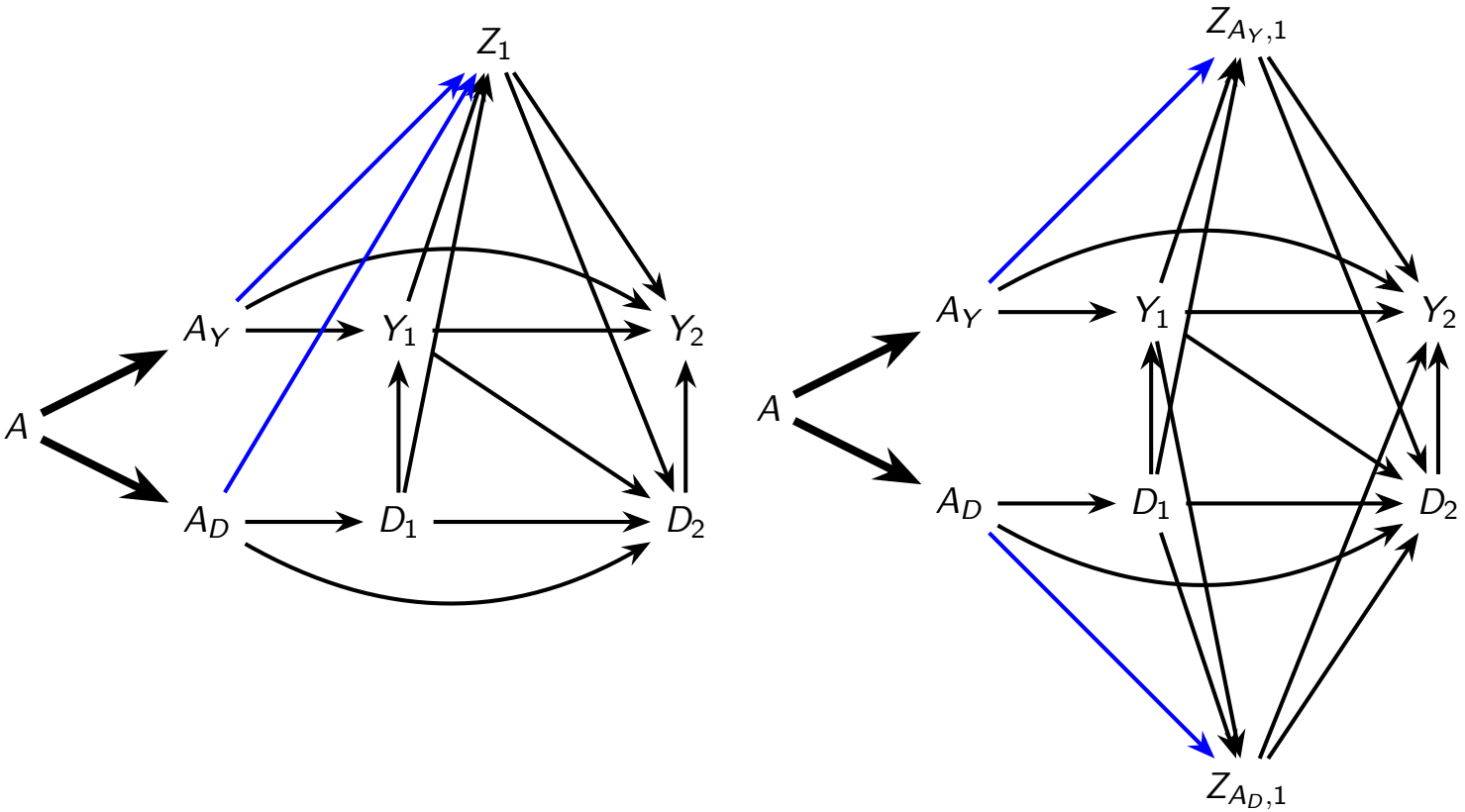
Extension to time-varying Z_k



The SPRINT example fits with the blue arrow in the left-hand graph.[4]

[4] Mats J Stensrud et al. "A generalized theory of separable effects in competing event settings". In: *Lifetime Data Analysis* (2021), pp. 1–44.

Extension to time-varying Z_k



Defining separable effects

- A_Y separable effects:

$$\Pr(Y_k^{a_Y=1,a_D}=1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D}=1), \quad a_D \in \{0,1\}$$

Defining separable effects

- A_Y separable effects:

$$\Pr(Y_k^{a_Y=1,a_D} = 1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D} = 1), \quad a_D \in \{0,1\}$$

~~*Effect of A on Y_k not through $\bar{D}_k$.*~~

Defining separable effects

- A_Y separable effects:

$$\Pr(Y_k^{a_Y=1,a_D}=1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D}=1), \quad a_D \in \{0,1\}$$

~~*Effect of A on Y_k not through $\bar{D}_k$.*~~

- A_D separable effects:

$$\Pr(Y_k^{a_Y,a_D=1}=1) \text{ vs. } \Pr(Y_k^{a_Y,a_D=0}=1), \quad a_Y \in \{0,1\}$$

Defining separable effects

- A_Y separable effects:

$$\Pr(Y_k^{a_Y=1,a_D}=1) \text{ vs. } \Pr(Y_k^{a_Y=0,a_D}=1), \quad a_D \in \{0,1\}$$

~~*Effect of A on Y_k not through $\bar{D}_k$.*~~

- A_D separable effects:

$$\Pr(Y_k^{a_Y,a_D=1}=1) \text{ vs. } \Pr(Y_k^{a_Y,a_D=0}=1), \quad a_Y \in \{0,1\}$$

~~*Effect of A on Y_k only through $\bar{D}_k$.*~~

Truncation event: Illustrative Example

VOLUME 24 · NUMBER 18 · JUNE 20 2006

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Quality of Life and Pain in Advanced Stage Prostate Cancer: Results of a Southwest Oncology Group Randomized Trial Comparing Docetaxel and Estramustine to Mitoxantrone and Prednisone

Donna L. Berry, Carol M. Moinpour, Caroline S. Jiang, Donna Pauler Ankerst, Daniel P. Petrylak, Lynne V. Vinson, Primo N. Lara, Sharon Jones, Mary E. Taplin, Patrick A. Burch, Maha H.A. Hussain, and E. David Crawford

Truncation event: Illustrative Example

VOLUME 24 · NUMBER 18 · JUNE 20 2006

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Quality of Life and Pain in Advanced Stage Prostate Cancer: Results of a Southwest Oncology Group Randomized Trial Comparing Docetaxel and Estramustine to Mitoxantrone and Prednisone

Donna L. Berry, Carol M. Moinpour, Caroline S. Jiang, Donna Pauler Ankerst, Daniel P. Petrylak, Lynne V. Vinson, Primo N. Lara, Sharon Jones, Mary E. Taplin, Patrick A. Burch, Maha H.A. Hussain, and E. David Crawford

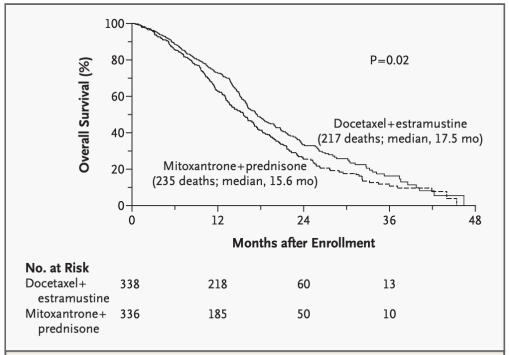


Figure 1. Kaplan–Meier Estimates of Overall Survival among Men with Androgen-Independent Prostate Cancer Treated with Mitoxantrone and Prednisone or Docetaxel and Estramustine.

Truncation event: Illustrative Example

VOLUME 24 · NUMBER 18 · JUNE 20 2006

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Quality of Life and Pain in Advanced Stage Prostate Cancer: Results of a Southwest Oncology Group Randomized Trial Comparing Docetaxel and Estramustine to Mitoxantrone and Prednisone

Donna L. Berry, Carol M. Moinpour, Caroline S. Jiang, Donna Pauler Ankerst, Daniel P. Petrylak, Lynne V. Vinson, Primo N. Lara, Sharon Jones, Mary E. Taplin, Patrick A. Burch, Maha H.A. Hussain, and E. David Crawford

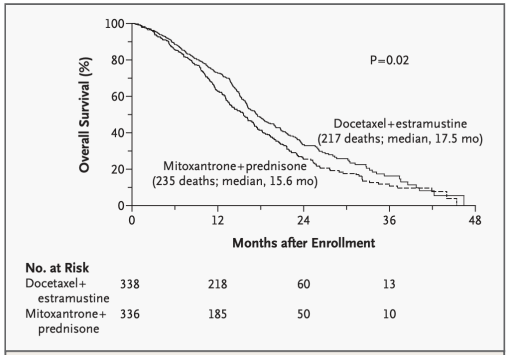


Figure 1. **Kaplan–Meier Estimates of Overall Survival among Men with Androgen-Independent Prostate Cancer Treated with Mitoxantrone and Prednisone or Docetaxel and Estramustine.**

Table 4. Mean Baseline and Change From Baseline QLQ-C30 Scores

		Baseline		Cycle 4		Cycle 8		1 Year	
Time	SD	DE	MP	DE	MP	DE	MP	DE	MP
Global QOL*		312	303	216	181	169	142	131	89
DE	22.3	60.6		+4.8		+5.1		−0.2	
MP	23.8	62.3		+6.9		+1.6		−2.7	

Truncation event: Illustrative Example

VOLUME 24 · NUMBER 18 · JUNE 20 2006

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Quality of Life and Pain in Advanced Stage Prostate Cancer: Results of a Southwest Oncology Group Randomized Trial Comparing Docetaxel and Estramustine to Mitoxantrone and Prednisone

Donna L. Berry, Carol M. Moinpour, Caroline S. Jiang, Donna Pauler Ankerst, Daniel P. Petrylak, Lynne V. Vinson, Primo N. Lara, Sharon Jones, Mary E. Taplin, Patrick A. Burch, Maha H.A. Hussain, and E. David Crawford

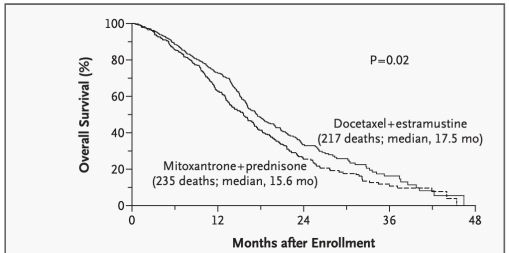


Figure 1. Kaplan–Meier Estimates of Overall Survival among Men with Androgen-Independent Prostate Cancer Treated with Mitoxantrone and Prednisone or Docetaxel and Estramustine.

Table 4. Mean Baseline and Change From Baseline QLQ-C30 Scores

		Baseline		Cycle 4		Cycle 8		1 Year	
Time	SD	DE	MP	DE	MP	DE	MP	DE	MP
Global QOL*		312	303	216	181	169	142	131	89
DE	22.3	60.6		+4.8		+5.1		−0.2	
MP	23.8	62.3		+6.9		+1.6		−2.7	

Several analogous settings (e.g. vaccine effects on post-infection outcomes).

This is not a COVID-19 talk, but...

One concern about COVID-19 vaccines is the antibody-dependent enhancement (ADE) phenomenon that vaccine could make the subsequent SARS-CoV-2 infection more severe.[26,27] The ADE phenomenon has been reported in studies of Middle East respiratory syndrome-CoV and SARS-CoV vaccines in animal challenge models.[26,27] However, this was not observed in the preclinical study in an immunization-challenge model of rhesus macaques using the same vaccines in the current study or in reports from preclinical studies of other COVID-19 vaccine candidates, including 2 other inactivated COVID-19 vaccines.[10,11,28] Studies have shown that previous infection of SARS-CoV-2 could protect against rechallenge in rhesus macaques.[29,30]

Xia et al, JAMA, 2020

Simple Notation

- Binary treatment $A \in \{0, 1\}$.
- Discrete time intervals $k = 0, 1, 2, ..., K + 1$.
- Event D_k (death from any cause).
- Covariates Z_k.
- Outcome of interest $Y \equiv Y_{K+1}$ (quality of life), which is undefined if $D_{K+1} = 1$.

Let's be explicit

- Let superscripts denote counterfactuals:

  Y^a is quality of life had, possibly contrary to fact, A been set to $a \in \{0, 1\}$.

  D_k^a is death by time k had, possibly contrary to fact, A been set to $a \in \{0, 1\}$.

How do we assess treatment effects on quality of life?

- Naive contrast:
  $\mathbb{E}(Y^{a=1} \mid D_K^{a=1} = 0)$ vs. $\mathbb{E}(Y^{a=0} \mid D_K^{a=0} = 0)$
- Intervene on death:
  $\mathbb{E}(Y^{a=1,d_K=0})$ vs. $\mathbb{E}(Y^{a=0,d_K=0})$
- The additive principal stratum effect:
  $\mathbb{E}(Y^{a=1} - Y^{a=0} \mid D_K^{a=1} = D_K^{a=0} = 0)$

  "This is a prime example of the pure mathematician's approach to a practical problem: stranded on a desert island with no food supplies but a crate of tinned tuna, he imagines a can opener"

  Dawid & Didelez (2012).

  See also a series of articles in International Journal of Biostatistics (2011), Robins (1986) and many more.

.....No other solution?

Let's take one step back (again)

Formal causal estimands are practically useful if they correspond to practically relevant questions.

JAMA and COVID

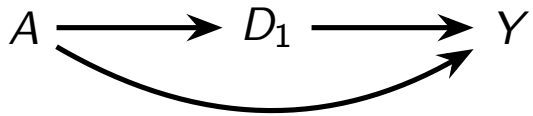
One concern about COVID-19 vaccines is the antibody-dependent enhancement (ADE) phenomenon that vaccine could make the subsequent SARS-CoV-2 infection more severe.[26,27] The ADE phenomenon has been reported in studies of Middle East respiratory syndrome–CoV and SARS-CoV vaccines in animal challenge models.[26,27] However, this was not observed in the preclinical study in an immunization-challenge model of rhesus macaques using the same vaccines in the current study or in reports from preclinical studies of other COVID-19 vaccine candidates, including 2 other inactivated COVID-19 vaccines.[10,11,28] Studies have shown that previous infection of SARS-CoV-2 could protect against rechallenge in rhesus macaques.[29,30]

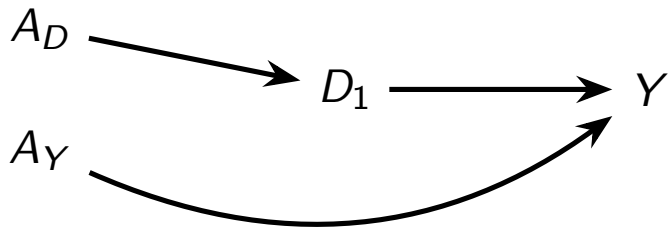
Another concern with whole-inactivated virus, particularly with alum adjuvant that can induce T helper 2 cell-biased responses, was vaccine-associated enhanced respiratory disease (VAERD). VAERD was reported in young children in the 1960s when whole-inactivated virus vaccine with alum adjuvant was tested for measles and respiratory syncytial virus.[31,32] However, most of the inactivated vaccines against COVID-19 under development used alum adjuvant,[10,11] and no evidence of VAERD has been seen. Instead, alum may reduce immunopathology compared with unadjuvanted coronavirus vaccines.[33] In the current study, notable changes in the lymphocyte subset distribution or various cytokines (including T helper 2 cell-related cytokines IL-4, IL-5, and IL-10) in various vaccine groups or alum-only group were not observed. However, T-cell–mediated immune responses on stimulation were not measured in the current study. Furthermore, alum is the most widely tested adjuvant component and has been commonly used in many types of vaccines on the market.[34] Nevertheless, safety, including the potential possibility of ADE and VAERD, will be closely monitored in the extended follow-up visits as well as in the phase 3 trial.

Xia et al, JAMA, 2020

A graphical condition: A_Y partial isolation

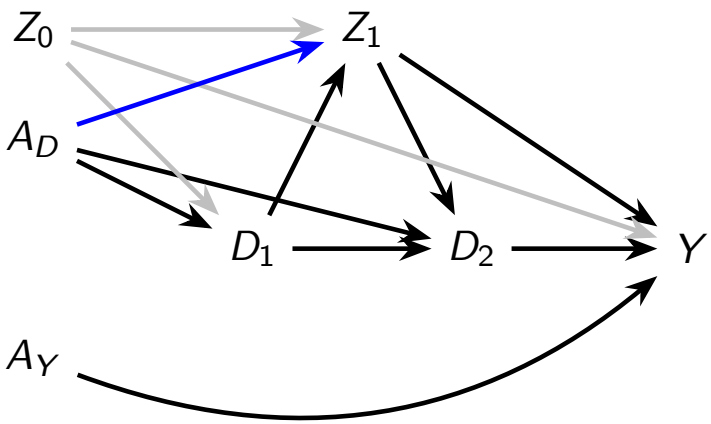
Suppose that

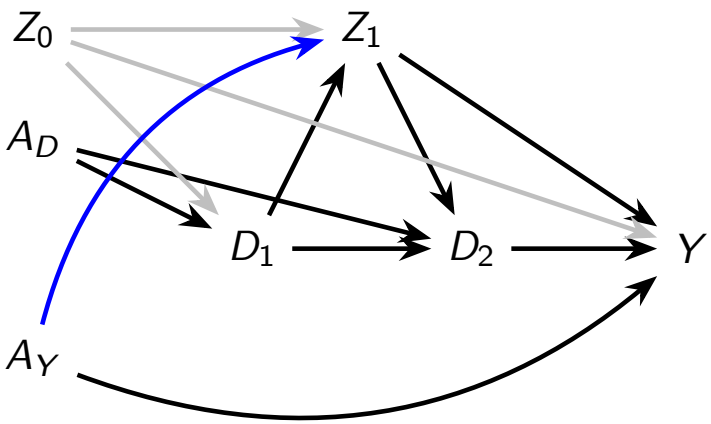




Inspired by Robins and Richardson (2010). See also Stensrud et al JASA (2020).

There are no causal paths from A_Y to D_{k+1}, $\quad k \in \{0, ..., K\}$.

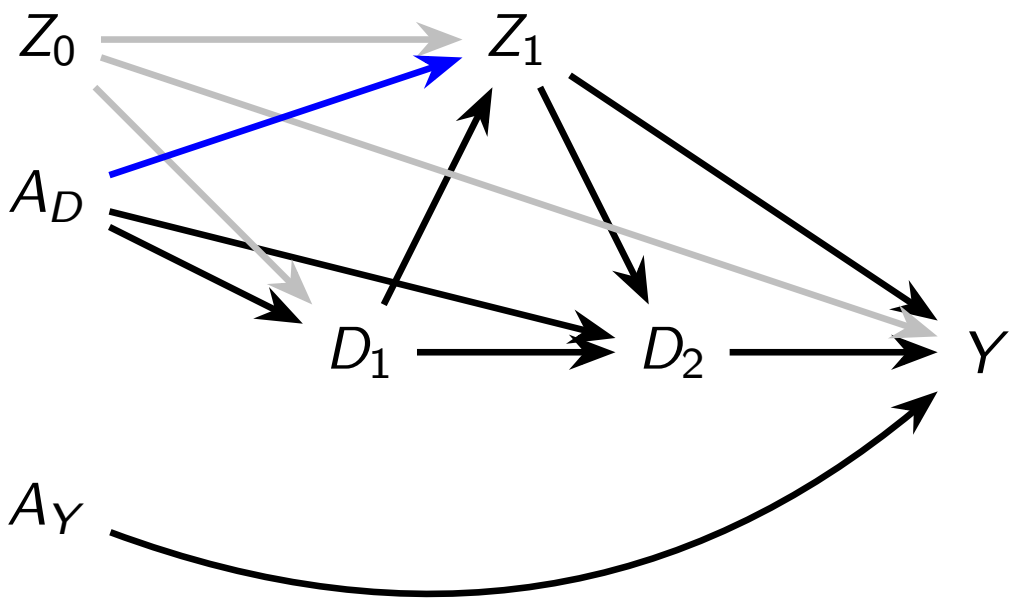




Conditional Separable Effects

"The effect of A on Y not through $\overline{D}_{K+1}$"

$$\mathbb{E}(Y^{a_Y=1,a_D} - Y^{a_Y=0,a_D} \mid D_K^{a_D} = 0), \quad a_D \in \{0,1\}$$
$$= \mathbb{E}(Y^{a_Y=1,a_D} - Y^{a_Y=0,a_D} \mid D_K^{a} = 0), \quad a \in \{0,1\}.$$



Relation to the Survivor Average Causal Effect (SACE)

- Suppose the effect of A on D_k is *monotone*: $D_k^{a=1} \leq D_k^{a=0}$.
- Then

$$\begin{aligned}
&\mathbb{E}(Y^{a=1} - Y^{a=0} \mid D_K^{a=1} = D_K^{a=0} = 0) \\
=&\mathbb{E}(Y^{a=1} - Y^{a=0} \mid D_K^{a=0} = 0) \quad \text{by monotonicity} \\
&\vdots \\
=&\mathbb{E}(Y^{a_Y=1,a_D=0} - Y^{a_Y=0,a_D=0} \mid D_K^{a_D=0} = 0)
\end{aligned}$$

$$\neq \mathbb{E}(Y^{a_Y=1,a_D=1} - Y^{a_Y=0,a_D=1} \mid D_K^{a_D=1} = 0).$$

which is our estimand of interest

Stories about modified treatments

Arguably, **stories about mechanisms are almost always about modified versions** of the observed treatment that change the mechanisms by which the treatment is hypothesized to act

- either by removing components of the treatment, or by
- adding on components that disable existing mechanisms.

Pearl's modified treatment story

Robins and Richardson[5] gave an example of this disconnect between stated causal effects and stories about clinical relevance in a **story told by Judea Pearl to defend the clinical relevance (actionable relevance) of natural effects** in a "usual" mediation setting.

[5] James M Robins and Thomas S Richardson. "Alternative graphical causal models and the identification of direct effects". In: *Causality and psychopathology: Finding the determinants of disorders and their cures* (2010), pp. 103–158.

Pearl's story about a 4 arm trial

Motivated a natural direct and indirect effects of smoking on MI outcomes relative to mediation by blood pressure.

- To defend this effect as "actionable" ("testable", "clinically relevant") he told a story about effects of modified cigarettes

Pearl's story about a 4 arm trial

Motivated a natural direct and indirect effects of smoking on MI outcomes relative to mediation by blood pressure.

- To defend this effect as "actionable" ("testable", "clinically relevant") he told a story about effects of modified cigarettes
- These could be tested in a 4-arm RCT with treatments
  1. Not smoke
  2. Smoke current cigarettes
  3. Smoke cigarettes with only nicotine removed
  4. Smoke cigarettes with only tobacco removed (e-cigarettes...)

The choice of estimands (our opinion)

- An explicit research question should always precede the choice of estimand: **it is the question that motivates the estimand, not the estimand that motivates the question**.[6]
- An implication of this message is that the choice of estimand critically depends on the context; **an estimand that is relevant in one study may be irrelevant in another.**

[6] Mats J Stensrud and Oliver Dukes. "Translating questions to estimands in randomized clinical trials with intercurrent events". In: *arXiv preprint arXiv:2111.08509* (2021).

Identification

estimand

Ingredients

- 150g unsalted butter, plus extra for greasing
- 150g plain chocolate, broken into pieces
- 150g plain flour
- ½ tsp baking powder
- ½ tsp bicarbonate of soda
- 200g light muscovado sugar
- 2 large eggs

Method

1. Heat the oven to 160C/140C fan/gas 3. Grease and base line a 1 litre heatproof glass pudding basin and a 450g loaf tin with baking parchment.

2. Put the butter and chocolate into a saucepan and melt over a low heat, stirring. When the chocolate has all melted remove from the heat.

estimator

estimate

Terminology

Remember the difference between the following terms:

- **Estimand** (a parameter of interest).
- **Estimator** (an algorithm / function that can be applied to data).
- **Estimate** (an output from applying the estimator to data).

We talk about bias of an estimator with respect to an estimand.
That is, the term *bias* (biased / unbiased) is defined with respect to an estimand.

Three tasks of data analyses that aim to answer practical questions (like, analyses of medical data)

1. Translation of the question into a formal causal estimand.
2. Assessing conditions for identification of the causal estimand.
3. Estimation of the causal estimand from our data.

Identification of Separable Effects

Requires identification of the quantities

$$\Pr(Y_{k+1}^{a_Y,a_D} = 1).$$

Problem: we observe only 2 out 4 treatment arms in the hypothetical trial of interest.

Identifiability conditions

- Exchangeability:

$$\bar{Y}^a_{K+1}, \bar{D}^a_{K+1} \perp\!\!\!\perp A \mid L \text{ for all } a.$$

- Consistency: if $A = a$, then

$$Y^a_{k+1} = Y_{k+1}$$
$$D^a_{k+1} = D_{k+1}.$$

- Positivity:

$$\begin{aligned}
&\Pr(L = l) > 0 \implies \\
&\quad \Pr(A = a \mid L = l) > 0 \text{ for } a \in \{0, 1\}, \\
&\Pr(D_{k+1} = Y_k = 0, L = l) > 0 \implies \\
&\quad \Pr(A = a | D_{k+1} = Y_k = 0, L = l) > 0 \\
&\text{for } a \in \{0, 1\} \text{ and } k \in \{0, \ldots, K\}.
\end{aligned}$$

Identifiability conditions

- Exchangeability:

$$\bar{Y}^a_{K+1}, \bar{D}^a_{K+1} \perp\!\!\!\perp A \mid L \text{ for all } a.$$

- Consistency: if $A = a$, then

$$Y^a_{k+1} = Y_{k+1}$$
$$D^a_{k+1} = D_{k+1}.$$

- Positivity:

$$\begin{aligned}
&\Pr(L = l) > 0 \implies \\
&\quad \Pr(A = a \mid L = l) > 0 \text{ for } a \in \{0, 1\}, \\
&\Pr(D_{k+1} = Y_k = 0, L = l) > 0 \implies \\
&\quad \Pr(A = a | D_{k+1} = Y_k = 0, L = l) > 0 \\
&\text{for } a \in \{0, 1\} \text{ and } k \in \{0, \ldots, K\}.
\end{aligned}$$

`Classical conditions. All of them, except the last positivity condition, hold in a perfectly executed RCT.`

Identification and estimation

Details are given in our manuscripts[7][8][9][10]

Take home message:

- Need to measure common causes of Y_k and D_k.
- no direct effect of A_D on Y_k and of A_Y on D_k.

[7] Stensrud et al., "A generalized theory of separable effects in competing event settings".

[8] Mats J Stensrud et al. "Separable effects for causal inference in the presence of competing events". In: *Journal of the American Statistical Association* (2020), pp. 1–9.

[9] Mats J Stensrud et al. "Conditional separable effects". In: *Journal of the American Statistical Association* just-accepted (2022), pp. 1–29.

[10] Torben Martinussen and Mats J Stensrud. "Estimation of separable direct and indirect effects in continuous time". In: *Biometrics* (2021).

Dismissible component conditions

$$\begin{aligned}\boldsymbol{\Delta 1}: &\Pr(Y_{k+1}^{a_Y, a_D=1} = 1 \mid Y_k^{a_Y, a_D=1} = 0, D_{k+1}^{a_Y, a_D=1} = 0, L = l) \\ &= \Pr(Y_{k+1}^{a_Y, a_D=0} = 1 \mid Y_k^{a_Y, a_D=0} = 0, D_{k+1}^{a_Y, a_D=0} = 0, L = l),\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Delta 2}: &\Pr(D_{k+1}^{a_Y=1, a_D} = 1 \mid Y_k^{a_Y=1, a_D} = 0, D_k^{a_Y=1, a_D} = 0, L = l) \\ &= \Pr(D_{k+1}^{a_Y=0, a_D} = 1 \mid Y_k^{a_Y=0, a_D} = 0, D_k^{a_Y=0, a_D} = 0, L = l),\end{aligned}$$

Dismissible component conditions

$$\begin{aligned}\boldsymbol{\Delta 1}: &\Pr(Y_{k+1}^{a_Y, a_D=1} = 1 \mid Y_k^{a_Y, a_D=1} = 0, D_{k+1}^{a_Y, a_D=1} = 0, L = l) \\ &= \Pr(Y_{k+1}^{a_Y, a_D=0} = 1 \mid Y_k^{a_Y, a_D=0} = 0, D_{k+1}^{a_Y, a_D=0} = 0, L = l),\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Delta 2}: &\Pr(D_{k+1}^{a_Y=1, a_D} = 1 \mid Y_k^{a_Y=1, a_D} = 0, D_k^{a_Y=1, a_D} = 0, L = l) \\ &= \Pr(D_{k+1}^{a_Y=0, a_D} = 1 \mid Y_k^{a_Y=0, a_D} = 0, D_k^{a_Y=0, a_D} = 0, L = l),\end{aligned}$$

These conditions may appear to be unintelligble, but

Dismissible component conditions

$$\begin{aligned}\boldsymbol{\Delta 1}: &\Pr(Y_{k+1}^{a_Y, a_D=1} = 1 \mid Y_k^{a_Y, a_D=1} = 0, D_{k+1}^{a_Y, a_D=1} = 0, L = l) \\ &= \Pr(Y_{k+1}^{a_Y, a_D=0} = 1 \mid Y_k^{a_Y, a_D=0} = 0, D_{k+1}^{a_Y, a_D=0} = 0, L = l),\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Delta 2}: &\Pr(D_{k+1}^{a_Y=1, a_D} = 1 \mid Y_k^{a_Y=1, a_D} = 0, D_k^{a_Y=1, a_D} = 0, L = l) \\ &= \Pr(D_{k+1}^{a_Y=0, a_D} = 1 \mid Y_k^{a_Y=0, a_D} = 0, D_k^{a_Y=0, a_D} = 0, L = l),\end{aligned}$$

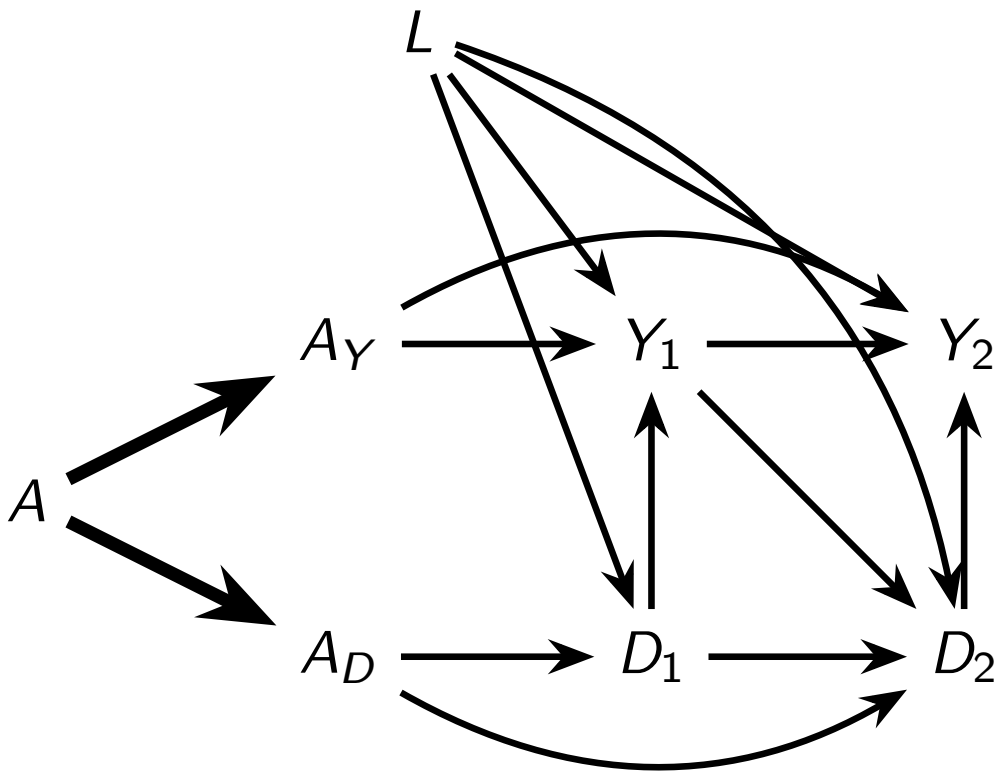
These conditions may appear to be unintelligble, but

- think about these conditions as equalities of hazards, and
- evaluate them in causal DAGs or SWIGs (next slides).

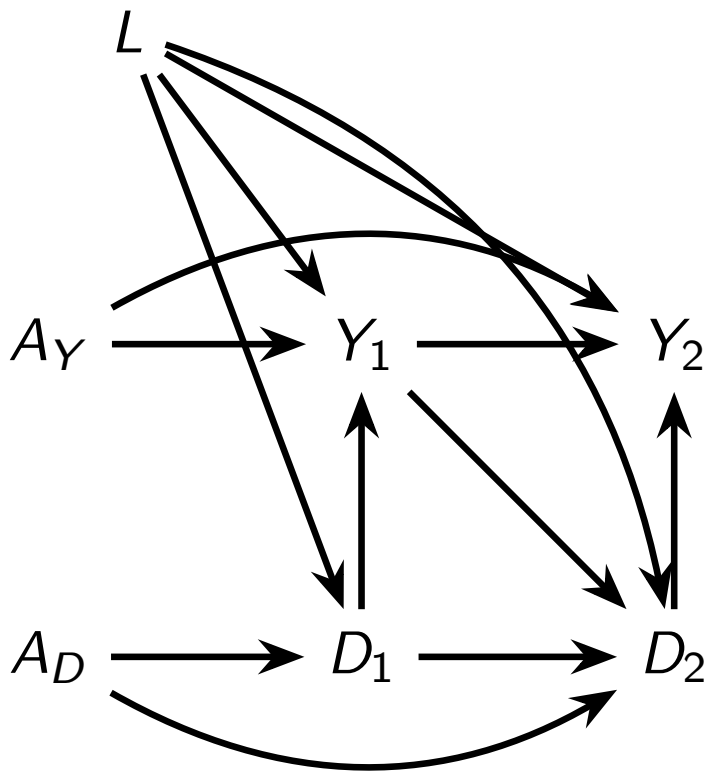
Again, the (informal) take home message is:

- Need to measure common causes of Y_k and D_k.
- no direct effect of A_D on Y_k and of A_Y on D_k.

DAG describing the observed data



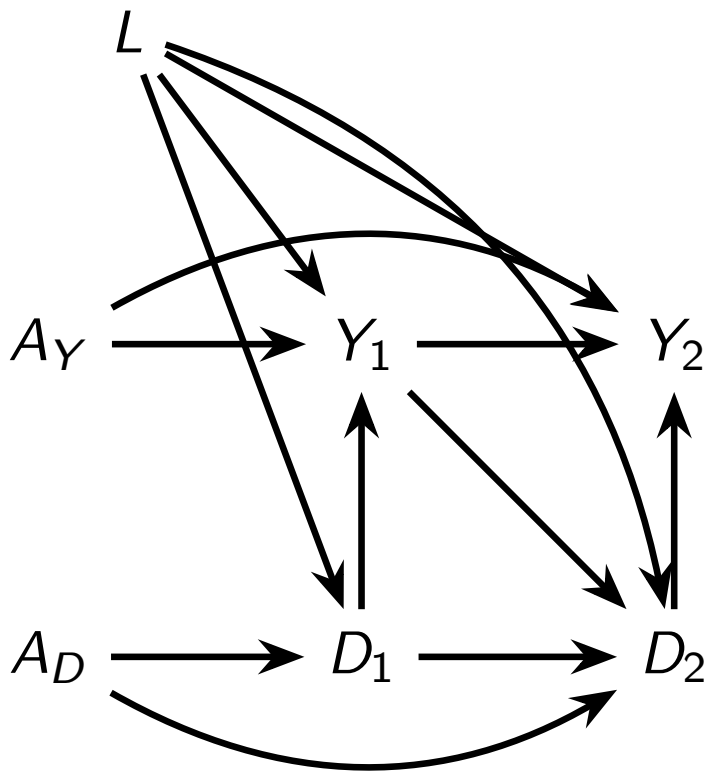
DAG describing hypothetical 4 arm trial G



(Bold arrows and A are removed)

$$\textbf{Δ1}: Y_{k+1}(G) \perp\!\!\!\perp A_D(G) \mid A_Y(G), Y_k(G) = 0, D_{k+1}(G) = 0, L(G),$$
$$\textbf{Δ2}: D_{k+1}(G) \perp\!\!\!\perp A_Y(G) \mid A_D(G), D_k(G) = 0, Y_k(G) = 0, L(G).$$

DAG that describes hypothetical 4 arm trial G

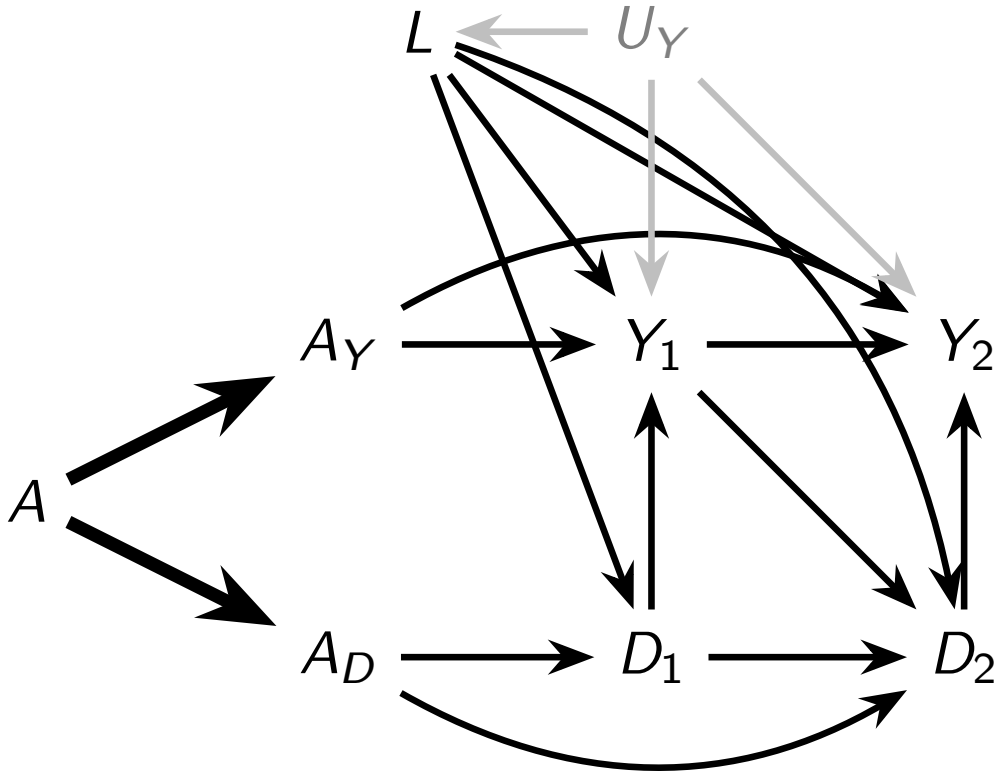


(Bold arrows and A are removed)

Δ1 : $Y_{k+1} \perp\!\!\!\perp A_D \mid A_Y, Y_k = 0, D_{k+1} = 0, L,$

Δ2 : $D_{k+1} \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, Y_k = 0, L.$

Understanding the dismissible component conditions

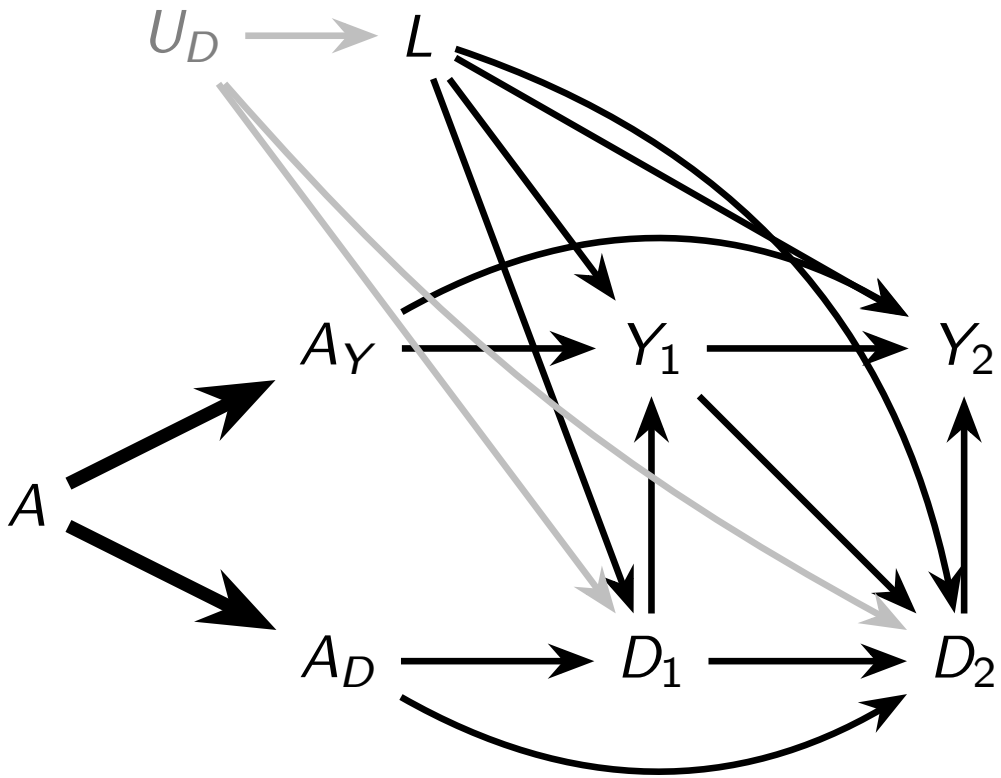


Conditions $\Delta 1$ and $\Delta 2$ hold.

$$\textbf{\Delta 1}: Y_{k+1} \perp\!\!\!\perp A_D \mid A_Y, Y_k = 0, D_{k+1} = 0, L,$$

$$\textbf{\Delta 2}: D_{k+1} \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, Y_k = 0, L.$$

Understanding the dismissible component conditions

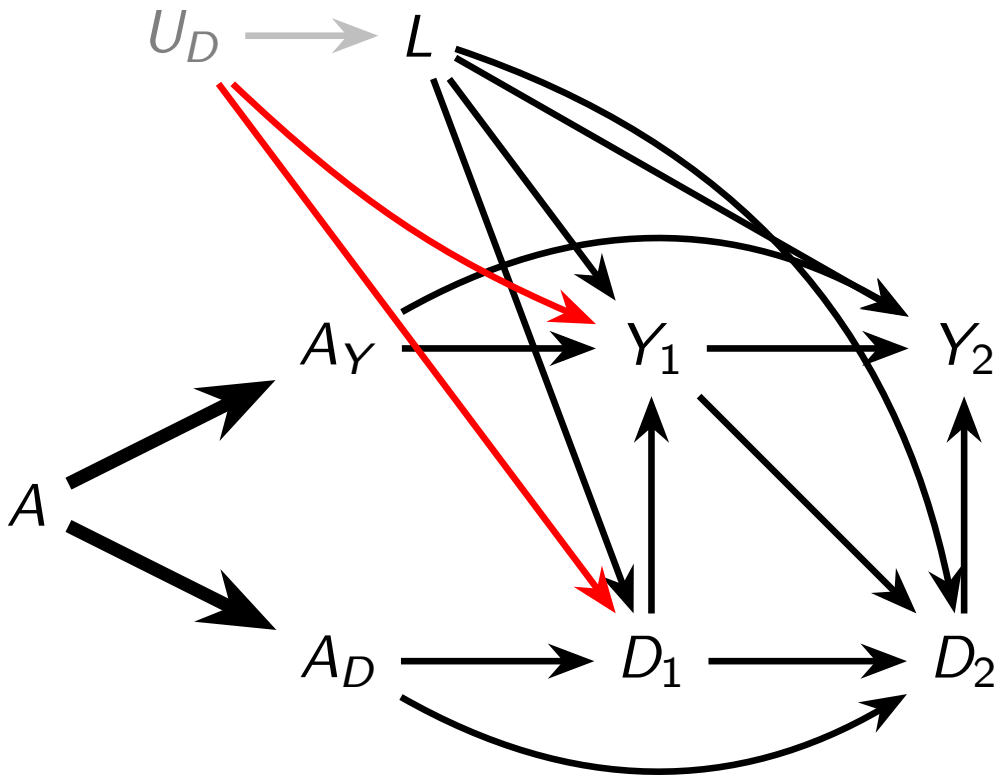


Conditions $\Delta 1$ and $\Delta 2$ hold.

$$\textbf{\Delta 1}: Y_{k+1} \perp\!\!\!\perp A_D \mid A_Y, Y_k = 0, D_{k+1} = 0, L,$$

$$\textbf{\Delta 2}: D_{k+1} \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, Y_k = 0, L.$$

Understanding the Dismissible Component Conditions



Conditions $\Delta 1$ is violated.

$$\textbf{\Delta 1}: Y_{k+1} \perp\!\!\!\perp A_D \mid A_Y, Y_k = 0, D_{k+1} = 0, L,$$

$$\textbf{\Delta 2}: D_{k+1} \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, Y_k = 0, L.$$

Identification formula for separable effects

$$\Pr(Y_{k+1}^{a_Y,a_D}=1)=\sum_l\Big[\sum_{s=0}^{K}\Pr(Y_{s+1}=1\mid D_{s+1}=Y_s=0,A=a_Y,L=l)$$
$$\prod_{j=0}^{s}\big[\Pr(D_{j+1}=0\mid D_j=Y_j=0,A=a_D,L=l)$$
$$\times\Pr(Y_j=0\mid D_j=Y_{j-1}=0,A=a_Y,L=l)\big]\Big]f(L=l),$$

This is a functional of hazard functions and the density of L.

Heuristic motivation for the identification formula

Let G be a (perfectly executed) trial where A_Y and A_D are randomly assigned, then

$$
\begin{aligned}
&\Pr(Y_{k+1}^{a_Y,a_D} = 1) \\
&= \Pr(Y_{k+1} = 1 \mid A_Y = a_Y, A_D = a_D) \\
&= \sum_l \Big[\sum_{s=0}^{k} \Pr(Y_{s+1} = 1 \mid D_{s+1} = Y_s = 0, A_Y = a_Y, A_D = a_D, L = l) \\
&\quad \prod_{j=0}^{s} \big[\Pr(D_{j+1} = 0 \mid D_j = Y_j = 0, A_Y = a_Y, A_D = a_D, L = l) \\
&\quad \times \Pr(Y_j = 0 \mid D_j = Y_{j-1} = 0, A_Y = a_Y, A_D = a_D, L = l) \big] \Big] f(L = l).
\end{aligned}
$$

Identification formula for separable effects

$$\Pr(Y_{k+1}^{a_Y,a_D}=1)=\sum_l\Big[\sum_{s=0}^{K}\Pr(Y_{s+1}=1\mid D_{s+1}=Y_s=0,A=a_Y,L=l)$$
$$\prod_{j=0}^{s}\big[\Pr(D_{j+1}=0\mid D_j=Y_j=0,A=a_D,L=l)$$
$$\times\Pr(Y_j=0\mid D_j=Y_{j-1}=0,A=a_Y,L=l)\big]\Big]f(L=l),$$

This is a functional of hazard functions and the density of L.
Algebraically equivalent weighted representations are described in Stensrud et al (2020), Journal of the American Statistical Association.

Identification formula for total effects (that Jessica discussed)

$$\Pr(Y_{k+1}^{a_Y,a_D}=1)=\sum_l\Big[\sum_{s=0}^{K}\Pr(Y_{s+1}=1\mid D_{s+1}=Y_s=0,A=a_Y,L=l)$$
$$\prod_{j=0}^{s}\big[\Pr(D_{j+1}=0\mid D_j=Y_j=0,A=a_Y,L=l)$$
$$\times\Pr(Y_j=0\mid D_j=Y_{j-1}=0,A=a_Y,L=l)\big]\Big]f(L=l),$$

This is a functional of hazard functions and the density of L.

Identification formula for controlled direct effects (that Jessica discussed)

$$\Pr(Y_{k+1}^{a_Y,a_D}=1)=\sum_l\Bigg[\sum_{s=0}^{K}\Pr(Y_{s+1}=1\mid D_{s+1}=Y_s=0,A=a_Y,L=l)$$
$$\prod_{j=0}^{s}\big[\Pr(D_{j+1}=0\mid D_j=Y_j=0,A=a_Y,L=l)$$
$$\times\Pr(Y_j=0\mid D_j=Y_{j-1}=0,A=a_Y,L=l)\big]\Bigg]f(L=l),$$

This is a functional of hazard functions and the density of L.

What does the identification formula mean to practitioners

Estimate the separable effect with continuous time models

$$\hat{\Pr}(Y_t^{a_Y=1,a}=1)-\hat{\Pr}(Y_t^{a_Y=0,a}=1),$$

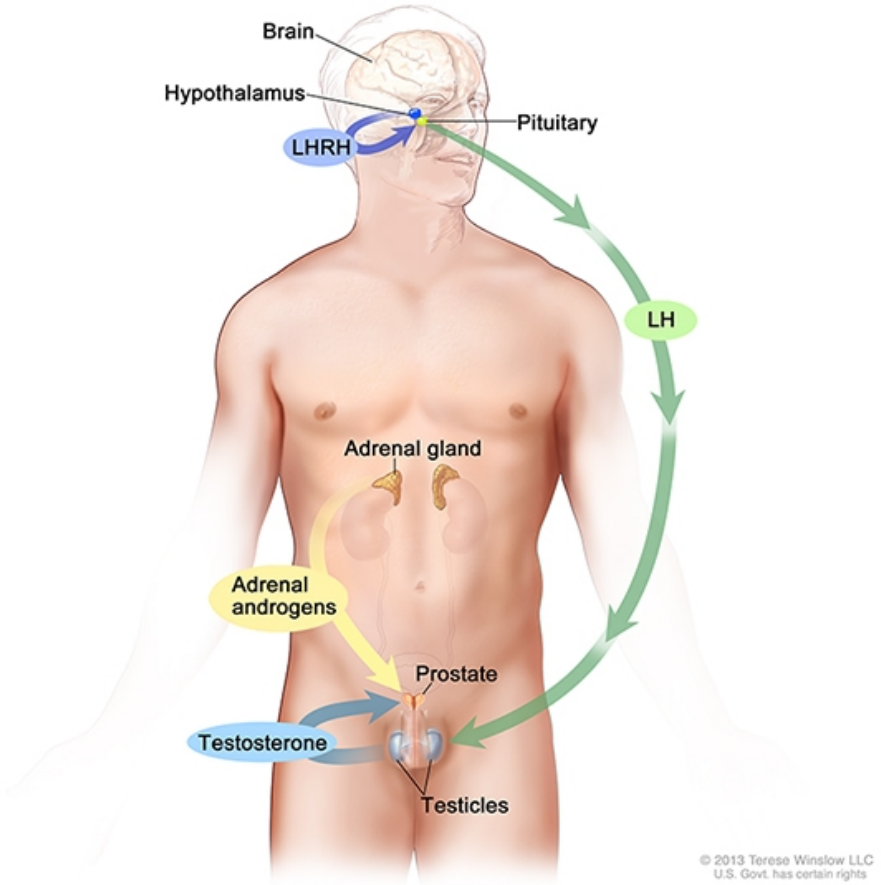
where

$$\hat{\Pr}(Y_t^{a_Y=1,a}=1)=n^{-1}\sum_i\left\{\int_0^t e^{-\hat{H}_1(s|A=1,L)-\hat{H}_2(s|A=a,L)}d\hat{H}_1(s|A=1,L)\right\},$$

$$\hat{\Pr}(Y_t^{a_Y=0,a}=1)=n^{-1}\sum_i\left\{\int_0^t e^{-\hat{H}_1(s|A=0,L)-\hat{H}_2(s|A=a,L)}d\hat{H}_1(s|A=0,L)\right\},$$

where $H_j(t|A=a,L)=\int_0^t h_j(s|A=a,L)\,ds$. because the terms in $\Pr(Y_t^{a_Y=1,a}=1)$ and $\Pr(Y_t^{a_Y=0,a}=1)$ can easily be estimated using Cox-models for the two cause specific hazard functions $h_j(s|A=a,L)$.

How does DES work?

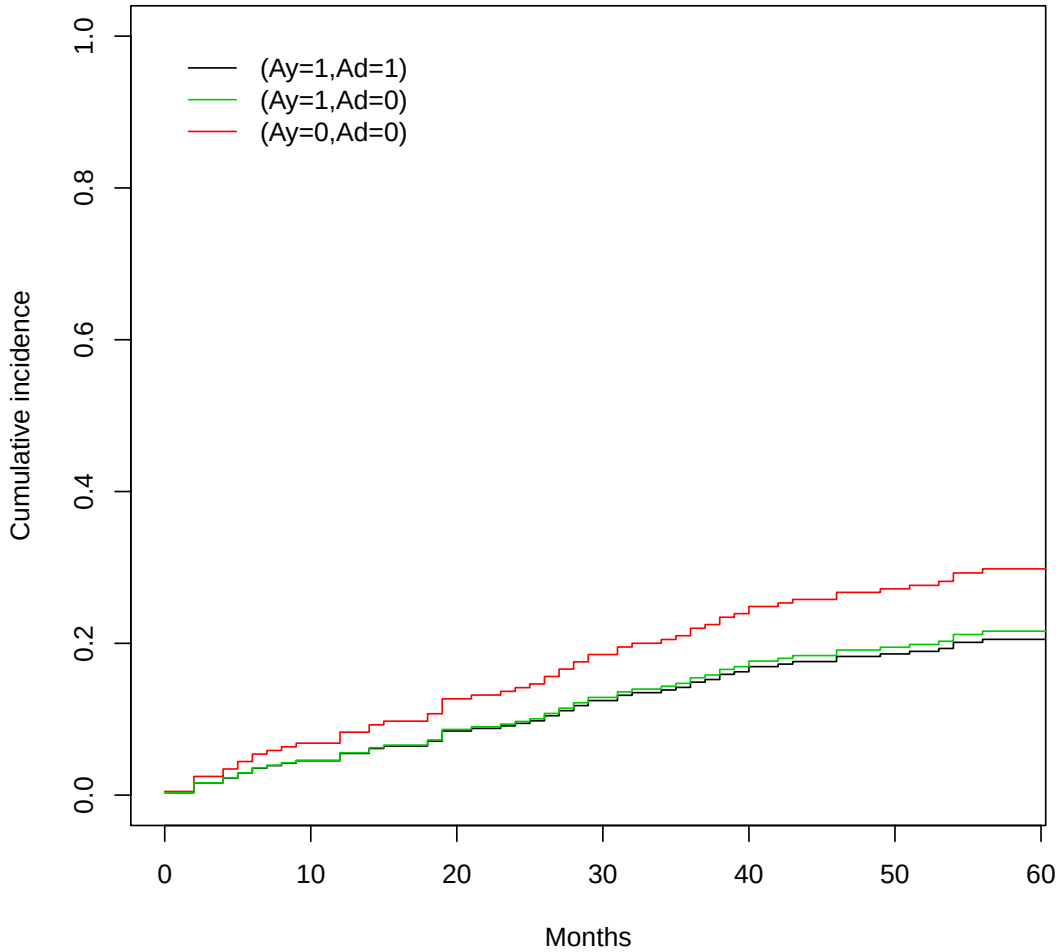


https://www.cancer.gov/types/prostate

Randomized Trial on Prostate Cancer Therapy (DES)

- A_Y : Exerts effects on death due to prostate cancer.
- A_D : Exerts effects on death due to other causes.
- Y_k : Death due to prostate cancer.
- D_k : Death due to other causes.

Randomised Trial on Prostate Cancer Therapy



Randomised Trial on Prostate Cancer Therapy

Table: Cumulative incidence after 3 years of follow-up.

Treatment	**Estimate (95%CI)**
$A_Y = 1, A_D = 1$	0.14 (0.08-0.20)
$A_Y = 1, A_D = 0$	0.15 (0.09-0.21)
$A_Y = 0, A_D = 0$	0.21 (0.15-0.28)

Previous identification formula

$$\Pr(Y_{k+1}^{a_Y,a_D}=1)=\sum_{l}\Big[\sum_{s=0}^{K}\Pr(Y_{s+1}=1\mid D_{s+1}=Y_s=0,A=a_Y,L=l)$$
$$\prod_{j=0}^{s}\big[\Pr(D_{j+1}=0\mid D_j=Y_j=0,A=a_D,L=l)$$
$$\times\Pr(Y_j=0\mid D_j=Y_{j-1}=0,A=a_Y,L=l)\big]\Big]f(L=l),$$

Generalized identification formula

$$\sum_{\bar{l}_k} \Big[\sum_{s=0}^{k} \Pr(Y_{s+1} = 1 \mid D_{s+1} = Y_s = 0, \bar{L}_s = \bar{l}_s, A = a_Y)$$

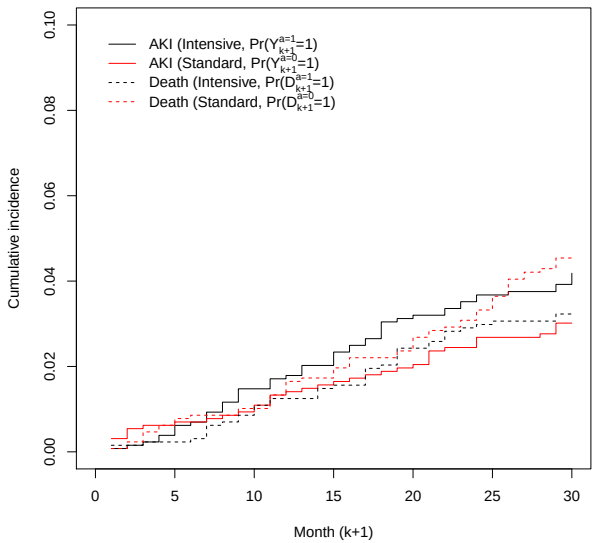
$$\prod_{j=0}^{s} \Big\{ \Pr(D_{j+1} = 0 \mid D_j = Y_j = 0, \bar{L}_j = \bar{l}_j, A = a_D)$$

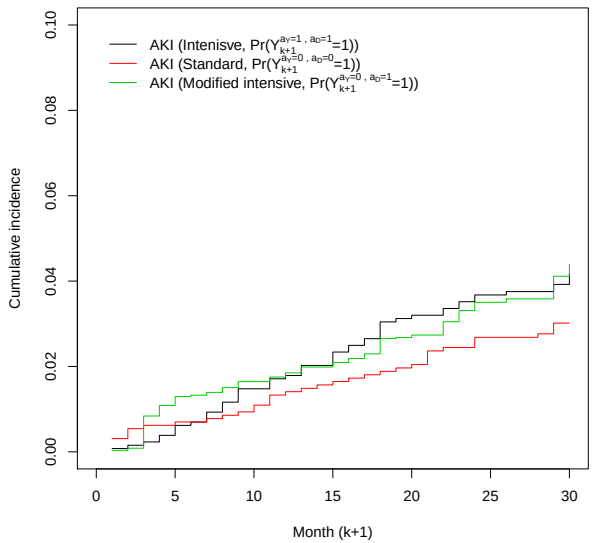
$$\times \Pr(Y_j = 0 \mid Y_{j-1} = D_j = 0, \bar{L}_{j-1} = \bar{l}_{j-1}, A = a_Y)$$

$$\times f(l_{A_Y,j} \mid Y_j = D_j = 0, \bar{l}_{j-1}, l_{A_D,j}, A = a_Y)$$

$$\times f(l_{A_D,j} \mid Y_j = D_j = 0, \bar{l}_{j-1}, A = a_D) \Big\} \Big].$$

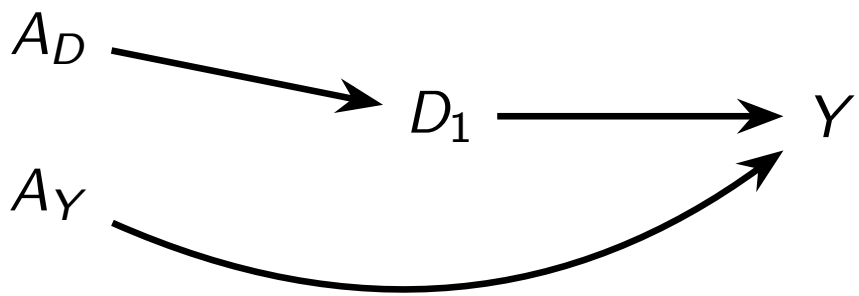
SPRINT example





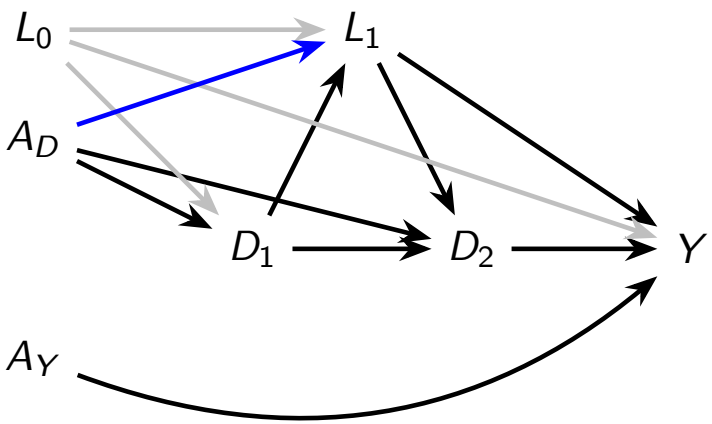
Dismissible Component Conditions

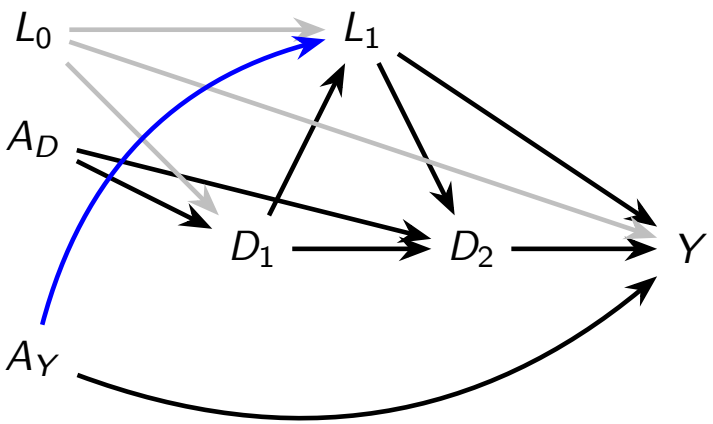
$$Y \perp\!\!\!\perp A_D \mid A_Y, D_{K+1} = 0, \overline{L}_K,$$
$$D_{k+1} \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, \overline{L}_k,$$
$$L_k \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, \overline{L}_{k-1}.$$



Dismissible Component Conditions

$$
\begin{aligned}
Y \perp\!\!\!\perp A_D \mid A_Y, D_{K+1} = 0, \overline{L}_K,\\
D_{k+1} \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, \overline{L}_k,\\
L_k \perp\!\!\!\perp A_Y \mid A_D, D_k = 0, \overline{L}_{k-1}.
\end{aligned}
$$





Identification formula for conditional separable effects

Simple g-formula:

$$\mathbb{E}(Y^{a_Y,a_D} \mid D_{K+1}^{a_Y,a_D} = 0) =$$
$$\frac{\sum_{\bar{l}_K} \mathbb{E}(Y \mid D_{K+1} = 0, \bar{L}_K = \bar{l}_K, A = a_Y) f_{\underline{L}_1, D_{K+1} \mid L_0, A}(\underline{l}_1, 0 \mid l_0, a_D) f_{L_0}(l_0)}{P(D_{K+1} = 0 \mid A = a_D)}.$$

Under monotonicity, this is also the identification formula for the principal stratum estimand (SACE).

Estimation: Simple outcome regression

Let $\hat{\nu}_{or,a_Y,a_D}$ be the solution to the estimating equation $\sum_{i=1}^{n} U_{or,i}(\nu_{a_Y,a_D}, \hat{\theta}) = 0$, where

$$U_{or,i}(\nu_{a_Y,a_D}, \hat{\theta}) = I(A_i = a_D)(1 - D_{k+1,i}) \times \left(\mathbb{E}(Y \mid D_{K+1} = 0, A = a_Y, \overline{L}_K; \hat{\theta}) - \nu_{a_Y,a_D} \right).$$

where $\hat{\theta}$ is the MLE of θ.

Estimation: Weighted estimator

Let $\hat{\nu}_{ipw,a_Y,a_D}$ be the solution to the estimating equation $\sum_{i=1}^{n} U_{ipw,i}(\nu_{a_Y,a_D}, \hat{\alpha}) = 0$, where

$$U_{ipw,i}(\nu_{a_Y,a_D}, \hat{\alpha}) = I(A_i = a_Y)(1 - D_{k+1,i})\hat{W}_i(a_Y, a_D; \hat{\alpha})\left(Y_i - \nu_{a_Y,a_D}\right),$$

and

$$\hat{W}(a_Y, a_D; \hat{\alpha}) = \frac{f_{\underline{L}_1,\overline{D}_{K+1}|L_0,A}(\underline{L}_1, 0 \mid L_0, a_D; \hat{\alpha})}{f_{\underline{L}_1,\overline{D}_{K+1}|L_0,A}(\underline{L}_1, 0 \mid L_0, a_Y; \hat{\alpha})}.$$

We also provide a doubly robust estimator.

Analysis of the prostate cancer trial

Estimand	Estimator	Estimate	(95% CI)
$\mathbb{E}(Y^{a=1} \mid D_{12}^{a=1} = 0)$	Non-param	-4.4	(-8.6, -0.5)
$\mathbb{E}(Y^{a=0} \mid D_{12}^{a=0} = 0)$	Non-param	-9.1	(-14.0, -3.9)
$\mathbb{E}(Y^{a_Y=0,a_D=1} \mid D_{12}^{a_Y=0,a_D=1} = 0)$	Ordinary reg	-6.1	(-11.8, -0.7)
$\mathbb{E}(Y^{a_Y=0,a_D=1} \mid D_{12}^{a_Y=0,a_D=1} = 0)$	IPW	-5.5	(-9.7, -0.4)
$\mathbb{E}(Y^{a_Y=0,a_D=1} \mid D_{12}^{a_Y=0,a_D=1} = 0)$	DR	-5.6	(-11.5, -0.2)

Table: Estimates of changes in health related quality of life at 12 months.

Summary

- Requires us to conceive hypothetical treatment decompositions.
  - *"This is a feature, not a bug"* (Thomas Richardson, 2020).
- Clarifies:
  - the notion of a conditional effect.
  - when a conditional effect is relevant for public health.
- A good exercise:
  - helps us think about future treatments.
- **No** cross-world independence assumptions.
- **No** hypothetical interventions to eliminate the competing event.

Extensions: Recurrent events

- Sometimes outcome of interest is not an indicator of *any* failure but *number* of events
- E.g. interest in treatment effect on number of fractures when some individuals die during follow-up
- Janvin et al. (forthcoming)

Some references

1. Stensrud MJ et al. *Separable Effects for Causal Inference in the Presence of Competing Risks*. Journal of the American Statistical Association (2020).
2. Stensrud MJ et al. *Generalized interpretation and identification of separable effects in competing risk settings* Lifetime data analysis and https://arxiv.org/abs/2004.14824 (2020)
3. Stensrud MJ et al. *Conditional separable effects* Journal of the American Statistical Association (2022)
4. Janvin M, Young JG, Ryalen PC, Stensrud MJ. *Causal inference with recurrent and competing events*. arXiv preprint arXiv:2202.08500. (2022)
5. Stensrud MJ, Dukes O. *Translating questions to estimands in randomized clinical trials with intercurrent events*. Statistics in Medicine (2022).
6. Sarvet AL, Stensrud MJ. *Without commitment to an ontology there could be no causal inference*. Epidemiology (2022).
7. Robins and Richardson. *Alternative graphical causal models and the identification of direct effects*. Causality and psychopathology: Finding the determinants of disorders and their cures (2010): 103-158.
8. Didelez V *Defining causal mediation with a longitudinal mediator and a survival outcome*. Lifetime data analysis (2018)
9. Young JG et al. *A causal framework for classical statistical estimands in failure time settings with competing events*. Statistics in Medicine (2020).
10. Robins JM et al. *An interventionist approach to mediation analysis* https://arxiv.org/abs/2008.06019

Some background on separable effects

- Didelez (2018)[11] extended this idea to settings with survival outcomes and time-varying mediator
- Stensrud et al. (2020)[12] identified this as a solution to the problem with the total effect in competing events settings.
- Now refer to this class of estimands as *separable effects*.
- Like all prior notions of effects that quantify mechanism, separable effects will always rely on stronger assumptions for identification than the total effect.
- But distinction is that, these effects inform real-world actions/decisions/policies.

[11] Vanessa Didelez. "Defining causal meditation with a longitudinal mediator and a survival outcome". In: *Lifetime data analysis* (2018), pp. 1–18.

[12] Stensrud et al., "Separable effects for causal inference in the presence of competing events".

Pure (natural) effects

The pure (natural) direct effect (Robins & Greenland; Pearl) is another definition of direct effect, when death is the relevant "mediator" this is

- $$\Pr[Y_{k+1}^{a=1,\overline{D}_{k+1}^{a=0}} = 1] \text{ vs. } \Pr[Y_{k+1}^{a=0,\overline{D}_{k+1}^{a=0}} = 1]$$

Indirect analogue together sums to total effect. Refers to an effect under intervention where the other death process is "set" for each individual to the value it would take under placebo. Similarly, unclear what this could inform in the real world.

Natural effects versus modified treatment effects

Robins and Richardson (2010)[13] posed a solution for the lack of actionable notions of direct effect which evolved from a story intended to argue the policy relevance of a natural (in)direct effect but instead told a story about different effects;

- effects of plausible modified versions of the study treatment under assumptions that these operate like the study treatment but with certain mechanisms removed.

Robins and Richardson (2010) posed assumptions under which this modified treatment effect would quantify (i) direct and indirect effects of the current treatment and (ii) be identified by the data in the current trial.

[13] Robins and Richardson, "Alternative graphical causal models and the identification of direct effects".

Hazards ratios are used all the time

DISCUSSION

For clinical trials of time-to-event outcomes, it has become standard practice to use Cox PH models both for trial design (e.g., power calculations) and statistical analysis. However, this may not be the best approach when the effect of treatment varies over time. Our analyses of 4 cardiology trials demonstrate some alternative approaches and outline some of their advantages and disadvantages under various patterns of treatment effect.

When PH are satisfied, the Cox PH model is the most statistically powerful method, and hazard ratios are readily understood by clinicians. We therefore see little practical reason to use alternative analysis strategies as the pre-specified primary analysis when deviation from PH is not expected, despite recent critiques of the hazard ratio for estimating treatment effects (13). However, when major deviations are

design beyond the choice of analysis strategy. When treatment is associated with a lower (or higher) short-term risk that later reverses, it is important that the trial continues for sufficient duration so that the long-term effects of the treatment can be fully understood. Longer-term (i.e., 5-year) results from the EXCEL study will therefore be helpful to further understanding of the risks and benefits of PCI relative to CABG in patients with left main coronary disease. A second implication for trial design is that the stopping criteria used by data monitoring committees should take into account potential non-PH patterns of treatment effect. For instance, the CHARM program Data Safety Monitoring Board did not recommend stopping early even though a planned interim analysis of short-term mortality showed a highly significant reduction in mortality on candesartan (17). Conversely, caution would be required when stopping a trial early for futility if a delayed effect was anticipated.

Gregson et al, JACC 2019.

`I disagree`[14]

[14] Mats J Stensrud and Miguel A Hernán. "Why test for proportional hazards?" In: *Jama* 323.14 (2020), pp. 1401–1402; Mats J Stensrud et al. "Limitations of hazard ratios in clinical trials". In: *European Heart Journal* (2018), ehy770.

Hazards ratios are used all the time

- Hazard ratios usually derived from a Cox model.

Hazards ratios are used all the time

- Hazard ratios usually derived from a Cox model.
- Relies on the proportional hazards assumption.

Hazards ratios are used all the time

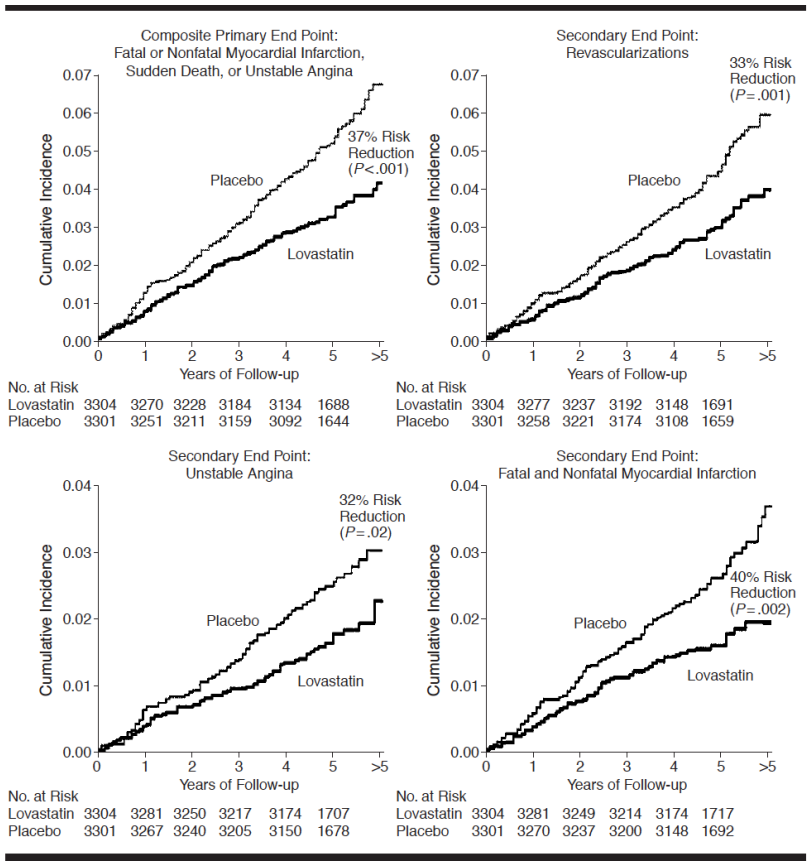
- Hazard ratios usually derived from a Cox model.
- Relies on the proportional hazards assumption.
- Researchers are asked to "test" whether hazards are proportional.

Hazards aren't expected to be proportional

Let's consider some paradigmatic settings,

- Statin therapy.
- Postmenopausal Estrogen.
- Cancer Screening.

Statins (JAMA)



- No effect until $\sim$ 9 months, then protective effect.

Downs et al, JAMA 1998

Postmenopausal Estrogen (NEJM)

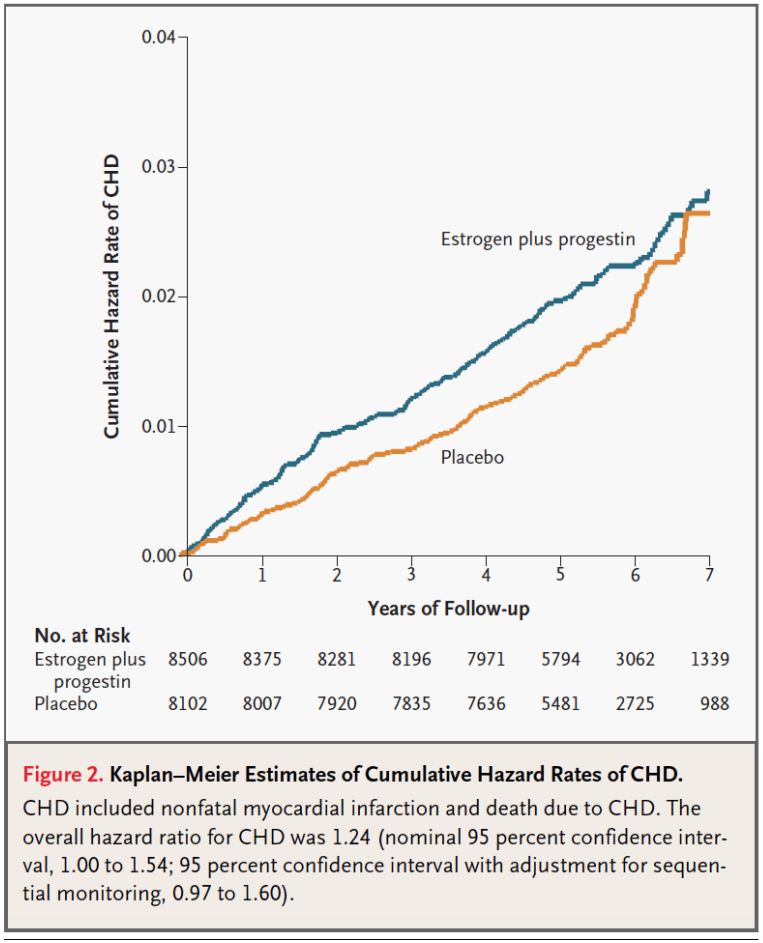
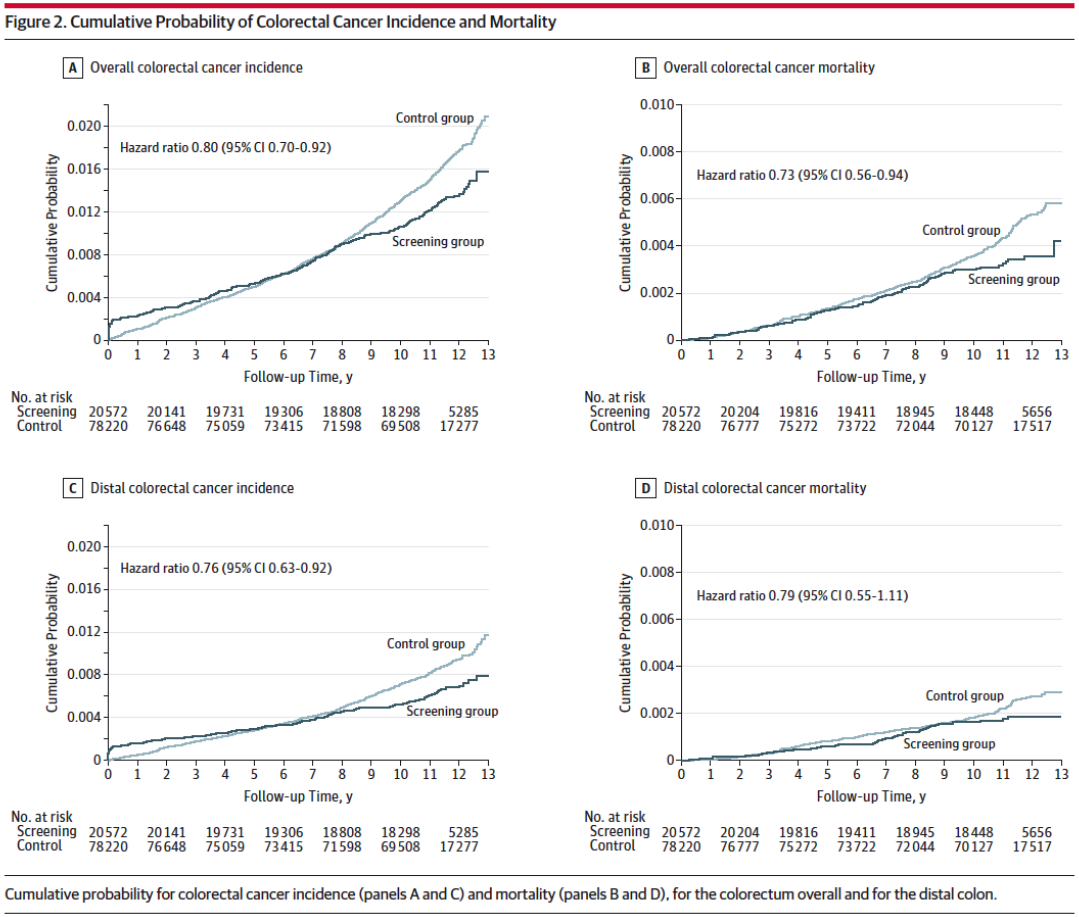


Figure 2. Kaplan–Meier Estimates of Cumulative Hazard Rates of CHD.
CHD included nonfatal myocardial infarction and death due to CHD. The overall hazard ratio for CHD was 1.24 (nominal 95 percent confidence interval, 1.00 to 1.54; 95 percent confidence interval with adjustment for sequential monitoring, 0.97 to 1.60).

- Clearly not proportional.

Manson et al, NEJM 2003

Colorectal cancer (JAMA)



- First harmful effect, then protective effect.

Holme et al, NEJM 2014

Hazards are not expected to be proportional

- The treatment effect usually changes over time.

Hazards are not expected to be proportional

- The treatment effect usually changes over time.
- Susceptibility varies across individuals.

Hazards are not expected to be proportional

- The treatment effect usually changes over time.
- Susceptibility varies across individuals.

  Greater susceptibility
  $\Longrightarrow$ more likely to develop the disease earlier.

Hazards are not expected to be proportional

- The treatment effect usually changes over time.
- Susceptibility varies across individuals.

  Greater susceptibility
  $\Longrightarrow$ more likely to develop the disease earlier.

  Remaining at risk is a proxy for being resilient to disease.

Hazards are not expected to be proportional

- The treatment effect usually changes over time.
- Susceptibility varies across individuals.

  Greater susceptibility
  $\Longrightarrow$ more likely to develop the disease earlier.

  Remaining at risk is a proxy for being resilient to disease.

  See e.g. Hernan et al (2010, Epidemiology).

Hazards are not expected to be proportional

- The treatment effect usually changes over time.
- Susceptibility varies across individuals.

  Greater susceptibility
  $\implies$ more likely to develop the disease earlier.

  Remaining at risk is a proxy for being resilient to disease.

  See e.g. Hernan et al (2010, Epidemiology).
- p-values > 0.05 are probably just due to underpowered tests.

How to interpret a Hazard Ratio from a Cox model

- Weighted average of time-varying hazard ratios.

Problems

- Variance estimates may be incorrect.

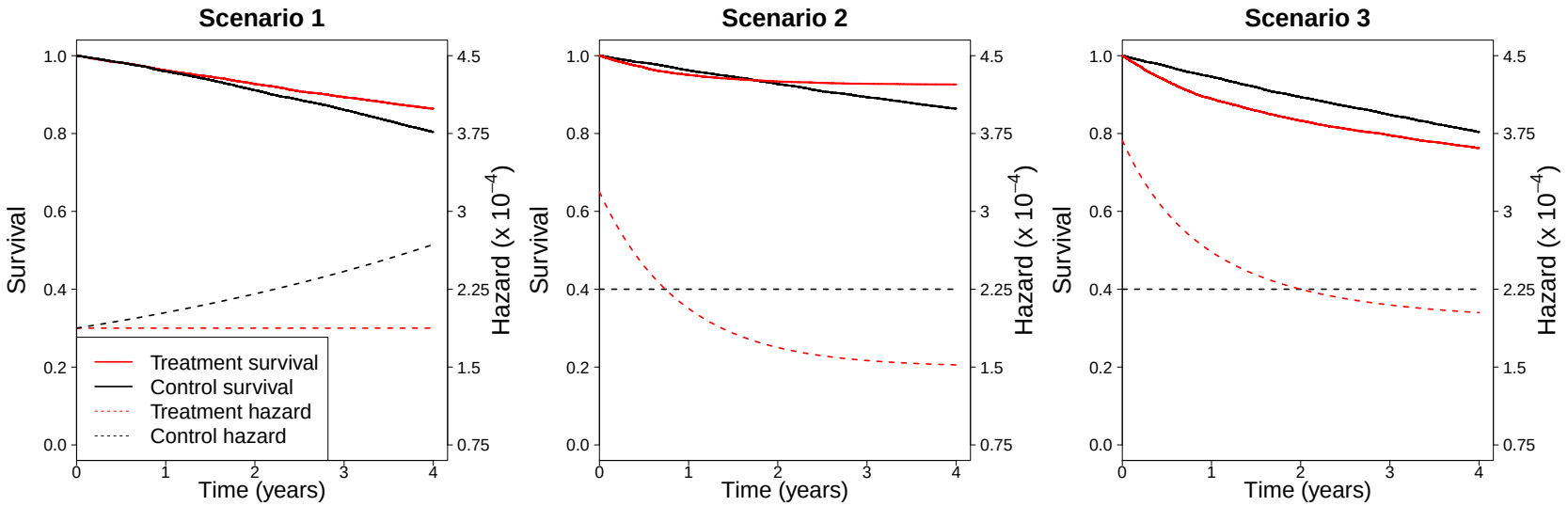
Problems

- Variance estimates may be incorrect.
- The magnitude of the hazard ratio depends on censoring

Problems

- Variance estimates may be incorrect.
- The magnitude of the hazard ratio depends on censoring even if the censoring is random.

Illustration of 3 trials



Mats J Stensrud and Miguel A Hernán. "Why test for proportional hazards?" In: *Jama* 323.14 (2020), pp. 1401–1402

The Magnitude of the HR depends on random censoring

Scenario	Censoring	Hazard ratio (95% CI), Cox proportional hazards model	4-year survival difference, % (95% CI), Kaplan-Meier estimator
1	No	0.68 (0.64, 0.72)	3.2% (2.4%, 4.0%)
	Yes	0.70 (0.65, 0.75)	3.4% (2.5%, 4.3%)
2	No	0.54 (0.50, 0.58)	3.5% (2.8%, 4.2%)
	Yes	0.66 (0.60, 0.72)	3.3% (2.5%, 4.1%)
3	No	1.28 (1.21, 1.35)	-5.3% (-6.2%, -4.3%)
	Yes	1.35 (1.27, 1.43)	-5.3% (-6.3%, -4.2%)

Additional references: Hazard ratios

1. Hernán MA. The hazards of hazard ratios.Epidemiology, 21(1):13–15, 2010
2. Hernán MA, Hernández-Díaz S, and Robins JM. A structural approach to selection bias. Epidemiology, 15:615–625, 2004
3. Martinussen T, Vansteelandt S, Andersen PK. Subtleties in the interpretation of hazard ratios. https://arxiv.org/abs/1810.09192
4. Stensrud MJ, Hernán MA. Why test for proportional hazards?. Jama. 2020 Apr 14;323(14):1401-2.
5. Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. European heart journal. 2019 May 1.

Some references

1. Robins, J., 1986. *A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect*. Mathematical modelling, 7(9-12), pp.1393-1512.
2. Robins and Richardson. *Alternative graphical causal models and the identification of direct effects*. Causality and psychopathology: Finding the determinants of disorders and their cures (2010): 103-158.
3. Didelez V *Defining causal mediation with a longitudinal mediator and a survival outcome*. Lifetime data analysis (2018)
4. Young JG et al. *A causal framework for classical statistical estimands in failure time settings with competing events*. Statistics in Medicine (2020).
5. Stensrud MJ et al. *Separable Effects for Causal Inference in the Presence of Competing Risks*. Journal of the American Statistical Association (2020).
6. Stensrud MJ et al. *Generalized interpretation and identification of separable effects in competing risk settings* Lifetime data analysis and https://arxiv.org/abs/2004.14824 (2020)
7. Stensrud MJ et al. *Conditional separable effects* https://arxiv.org/abs/2006.15681 (2020)
8. Janvin M, Young JG, Ryalen PC, Stensrud MJ. Causal inference with recurrent and competing events. arXiv preprint arXiv:2202.08500. (2022)
9. Stensrud MJ, Dukes O. Translating questions to estimands in randomized clinical trials with intercurrent events. arXiv preprint arXiv:2111.08509.
10. Robins JM et al. *An interventionist approach to mediation analysis* https://arxiv.org/abs/2008.06019

Additional references: Hazard ratios

1. Hernán MA. The hazards of hazard ratios.Epidemiology, 21(1):13–15, 2010
2. Hernán MA, Hernández-Díaz S, and Robins JM. A structural approach to selection bias. Epidemiology, 15:615–625, 2004
3. Martinussen T, Vansteelandt S, Andersen PK. Subtleties in the interpretation of hazard ratios. https://arxiv.org/abs/1810.09192
4. Stensrud MJ, Hernán MA. Why test for proportional hazards?. Jama. 2020 Apr 14;323(14):1401-2.
5. Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. European heart journal. 2019 May 1.