Causal inference with competing events

Identification and censoring

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

SER 2023, Portland

Return to the "causal road map"

The causal road map:

1. linking a particular causal effect we want to learn about
2. to a particular statistical method that we can apply to a data set
3. under explicit assumptions that we can reason through with subject matter (underlying causal model)
   - Causal diagrams (DAGs, SWIGs) useful tools for communicating subject matter assumptions
   - These types of assumptions formally called "identifying assumptions" (e.g. no unmeasured confounding)

We're now going to talk more about the "road trip" / "journey", once we've selected our target.

Focus will be on *total effects* and *controlled direct effects* because these are counterfactual contrasts in popular targets from the survival analysis literature

- *cause-specific cumulative incidence* and the *marginal cumulative incidence*

which can be estimated using well-studied and described survival analysis methods for "censored data".

- We'll use our "causal road map" to motivate these methods to give us some insight about when and how to apply them in competing event settings.

Identifying the total effect in the idealized trial

We said before that subject matter assumptions for identifying the total effect of A on Y_k hold by design in the idealized trial, with no loss to follow-up

- we have complete knowledge of the D and Y event process up to the administrative study end (e.g. for all $k = 1, \ldots, 365 \times 5$)

This means that we can write the total effect in this case $\Pr[Y_k^{a=1} = 1]$ vs. $\Pr[Y_k^{a=0} = 1]$ as the simple function of measured factuals:

$$\Pr[Y_k = 1|A = 1] \text{ vs. } \Pr[Y_k = 1|A = 0]$$

Why?

Randomization and identification

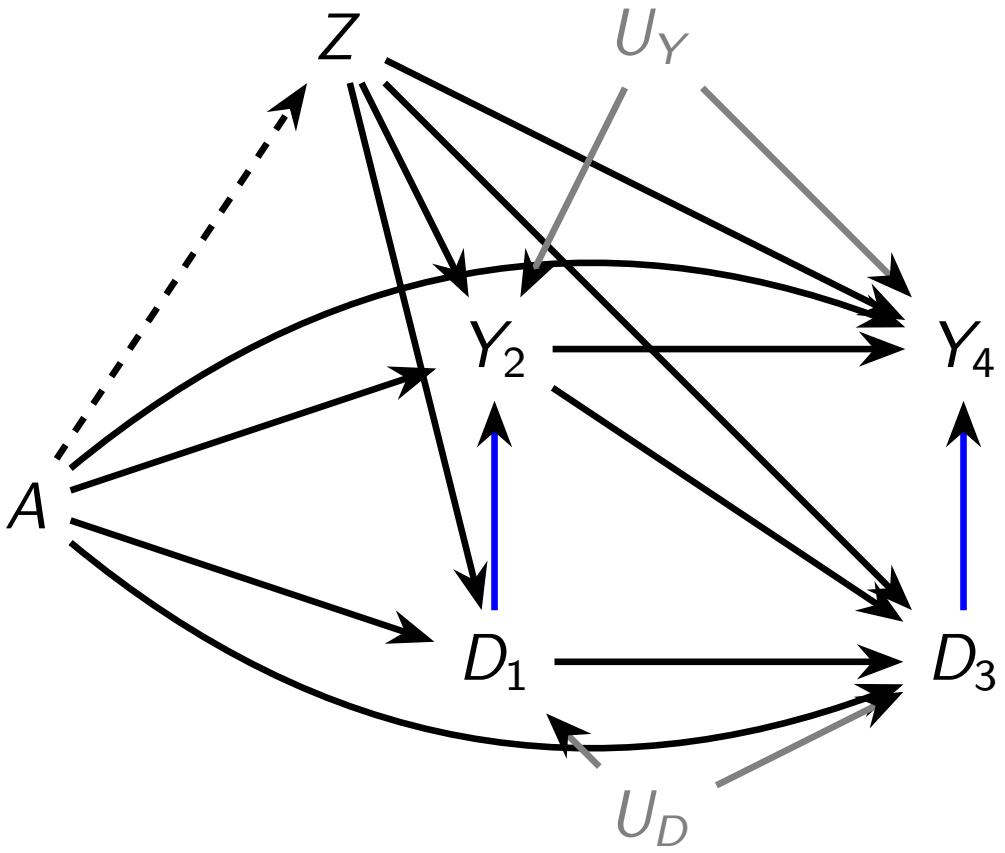
Because the value of A is determined only by the coin flip, we have that

$$Y_k^a \coprod A$$

where $\coprod$ denotes independence for either $a = 1$ or $a = 0$. In other words, the value of A an individual gets is not associated with their future counterfactual outcomes unconditionally (a.k.a. marginally) with respect to any covariates.

- This independence is an assumption but it is guaranteed to hold by our physical randomized assignment of A.
- This assumption is sometimes called "marginal exchangeability" or "no confounding"

Reading "no confounding" from causal DAG



No "backdoor paths" connecting A and future Y

No loss to follow-up

Randomization of A allows us to equate the total effect $\Pr[Y_k^{a=1} = 1]$ vs. $\Pr[Y_k^{a=0} = 1]$ to the function of factuals

$$\Pr[Y_k = 1|A = 1] \text{ vs. } \Pr[Y_k = 1|A = 0]$$

And then no loss to follow-up, means we can just estimate this with sample proportions since there is no missingness in Y_k for any k.

Identification in "censored" data

Real world studies are rarely "ideal" even when we just restrict ourselves to intention to treat effects in a trial (an effect of baseline randomization)

- One common issue is "loss to follow-up" which is a type of "censoring" event

Next we are going to think about assumptions we would need to identify the total and controlled direct effect in a trial where there is the complication of loss to follow-up.

- We will also consider conditions for classifying other events as "censoring events" and how this impacts assumptions for identification.

Less idealized estrogen therapy trial – notation

Consider a less "idealized" version of this estrogen therapy trial where individuals assigned to treatment $A = 1$ or $A = 0$ such that on every day $k = 1, \ldots, K$ (e.g. $K = 365 \times 5$ days),

- C_k indicator of "loss to follow-up" by k. Meaning, an event occurred by k prior to failure from any cause resulting in the whole future becoming missing (unknown) to investigators.
- When $C_k = 0$ we know $\overline{Y}_k$ indicator of event of interest *by* k (e.g. prostate cancer death) and $\overline{D}_k$ indicator of competing event *by* k (e.g. other death.).
- But when $C_k = 1$, the future $\underline{Y}_k, \underline{D}_k$ is missing.

Suppose we also thought to measure some covariates

Also suppose while alive and under observation, we are able to measure time-updated health characteristics in each interval k (L_k)

- e.g. new diagnoses, medications

As well as pre-randomization characteristics (e.g. age, race/ethnicity, family medical history), denote these L_0

Crucial difference between loss to follow-up and a competing event in the real world

$C_k = 1$

- if John is lost to follow-up on day 10 of the study period and he was alive as of day 9, we don't know if and when he later died of prostate cancer or anything else.

$D_k = 1$

- if John dies of a heart attack on day 10 of the study period and he was alive as of day 9, we know he never dies of prostate cancer.

Competing events versus censoring events

Competing events are routinely equated with "censoring events".

- We now give a general definition of a "censoring event" that helps to clarify when competing events should be classified as censoring events and when they should not.

This definition will lead us to derive the classical Aalen Johansen cause-specific cumulative incidence estimator and the Kaplan-Meier survival estimator using the steps of the "causal road map".

A general definition of a censoring event

A censoring event is any event occurring in the study that ensures the values of all future **counterfactual outcomes under treatment level** *a* **that are of interest are unknown/missing even for an individual who actually received treatment level** *a*

- Thus, what constitutes a censoring event depends on what is of interest (what our question is, what constitutes the top of the "road map")
- For now, let's generically denote an indicator of experiencing any censoring event by k, V_k

Censoring for the total effect

We defined the total effect at time k as

$$\Pr[Y_k^{a=1} = 1] \text{ vs. } \Pr[Y_k^{a=0} = 1]$$

Thus, if our question is about a total effect, the counterfactual outcomes of interest under a are Y_k^a. Even for an individual who received treatment level a, this outcome will be unknown/missing for an individual who is lost to follow-up prior to k.

- So for the total effect loss to follow-up is a censoring event, $V_k = C_k$, but competing events are NOT censoring events.

Censoring for the controlled direct effect

We defined the controlled direct effect as

$$\Pr[Y_k^{a=1,\bar{d}=0} = 1] \text{ vs. } \Pr[Y_k^{a=0,\bar{d}=0} = 1]$$

Thus, if our question is about a controlled direct effect, the counterfactual outcomes of interest under a are $Y_k^{a,\bar{d}=0}$. Even for an individual who received treatment level a, this outcome will be unknown/missing for an individual who **either** is lost to follow-up OR experiences the competing event prior to k.

- For the controlled direct effect BOTH loss to follow-up and competing events are censoring events, $V_k = (C_k, D_k)$

Censoring and identification

When there are censoring events in the study, we will need assumptions beyond those guaranteed by randomization of A to identify the effect of interest.

- We *may* be able to identify this effect *if* we can justify an additional exchangeability assumption with respect to the censoring event(s) V_k

Conditional exchangeability for censoring

Consider the following counterfactual independence assumption for all $k = 1, \ldots, K-1$:

$$Y_K^{a,\overline{v}=0}, \ldots, Y_k^{a,\overline{v}=0} \coprod V_k | \overline{L}_{k-1}, \overline{V}_{k-1} = \overline{Y}_{k-1} = 0, A = a$$

This is a type of *conditional (sequential) exchangeability* and is sometimes called *no unmeasured selection bias*

- Sometimes refer to the measured covariate history $\overline{L}_{k-1}$ as the *measured selection factor history for censoring*
- To make this hold we need to have in $\overline{L}_{k-1}$ all the *common causes* (or sufficient "proxies") of censoring V_k and the future event of interest process Y.

Competing events as either censoring events or "selection factors" for censoring

$$Y_K^{a,\overline{v}=0}, \ldots, Y_k^{a,\overline{v}=0} \coprod V_k | \overline{L}_{k-1}, \overline{V}_{k-1} = \overline{Y}_{k-1} = 0, A = a$$

How competing events play into this assumption depends on what is a censoring event

- When we are targeting a **controlled direct effect**, we need to consider this assumption with competing events a component of the vector of censoring events V_k
- When we are targeting a **total effect**, with C the only form of censoring, we need to consider this assumption with competing events an implicit component of the "selection factors for censoring events" $\overline{L}_{k-1}$

Note

I've written the exchangeability condition generally on in terms of counterfactuals $Y_k^{a,\overline{v}=0}$ indexed by an intervention $\overline{v} = 0$ or "eliminate censoring".

- This would suggest that all counterfactuals we consider include "elimination of loss to follow-up".
- The identification argument does rely on this meaning of the counterfactuals in censored data where loss to follow-up occurs.
- To claim identification for counterfactual target without this interpretation we could make an additional assumption that "C does not cause Y" such that, e.g. $Y^a = Y^{a,\overline{c}=0}$ for all individuals

I drop the $\overline{c} = 0$ index from all counterfactuals subsequently referenced which is making this assumption implicitly.

Competing events as censoring events

Meaning interest is in a controlled direct effect:

$$Y_K^{a,\overline{d}=0}, \ldots, Y_k^{a,\overline{d}=0} \coprod (C_k, D_k) | \overline{L}_{k-1}, \overline{C}_{k-1} = \overline{D}_{k-1} = \overline{Y}_k = 0, A = a$$

Here $\overline{L}_{k-1}$ needs to contain all common causes (or sufficient "proxies") of C and all future Y, **AND** all common causes (or sufficient "proxies") of D and all future Y.

Competing events as "selection factors" for censoring

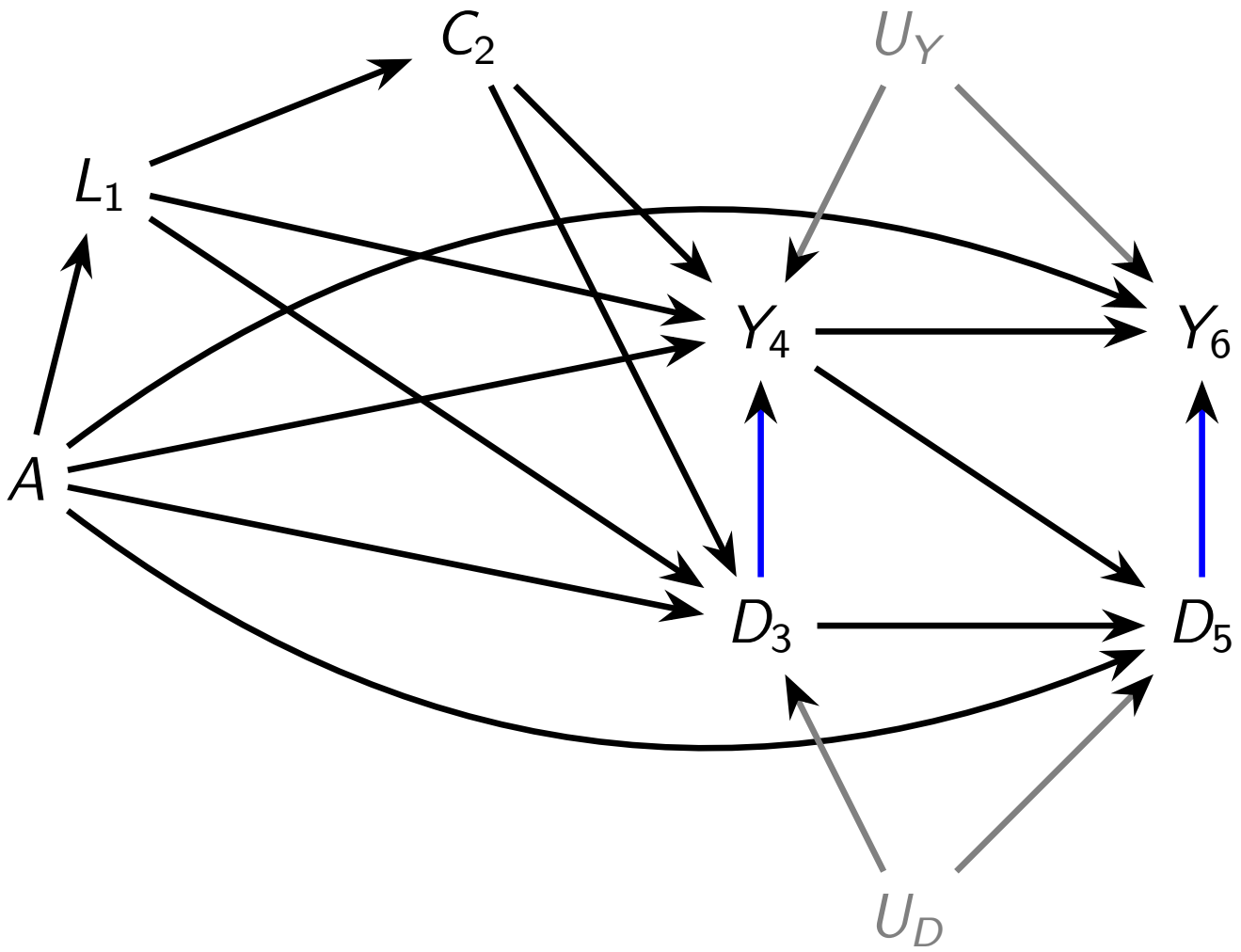
Meaning interest is in a total effect:

$$Y_K^a, \ldots, Y_k^a \coprod C_k | \overline{L}_{k-1}, \overline{D}_{k-1} = \overline{C}_{k-1} = \overline{Y}_{k-1} = 0, A = a$$

Here $\overline{L}_{k-1}$ needs to **only** contain all common causes (or sufficient "proxies") of C and future Y. D here is not a censoring event but will ALWAYS be a common cause of future C (censoring in this case) and Y.

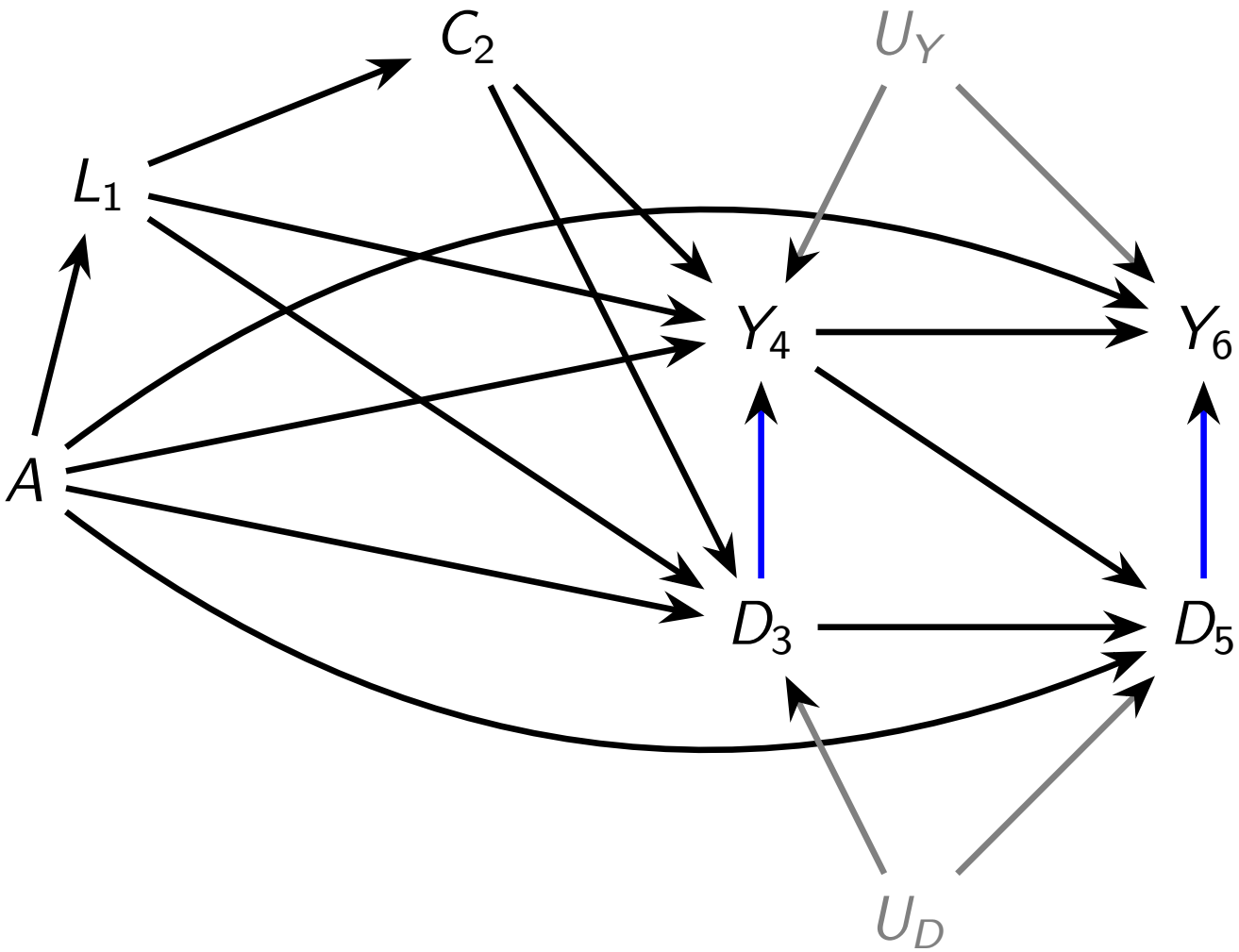
- Why is D always a common cause of C and Y?

Causal DAG for study with loss to follow-up.



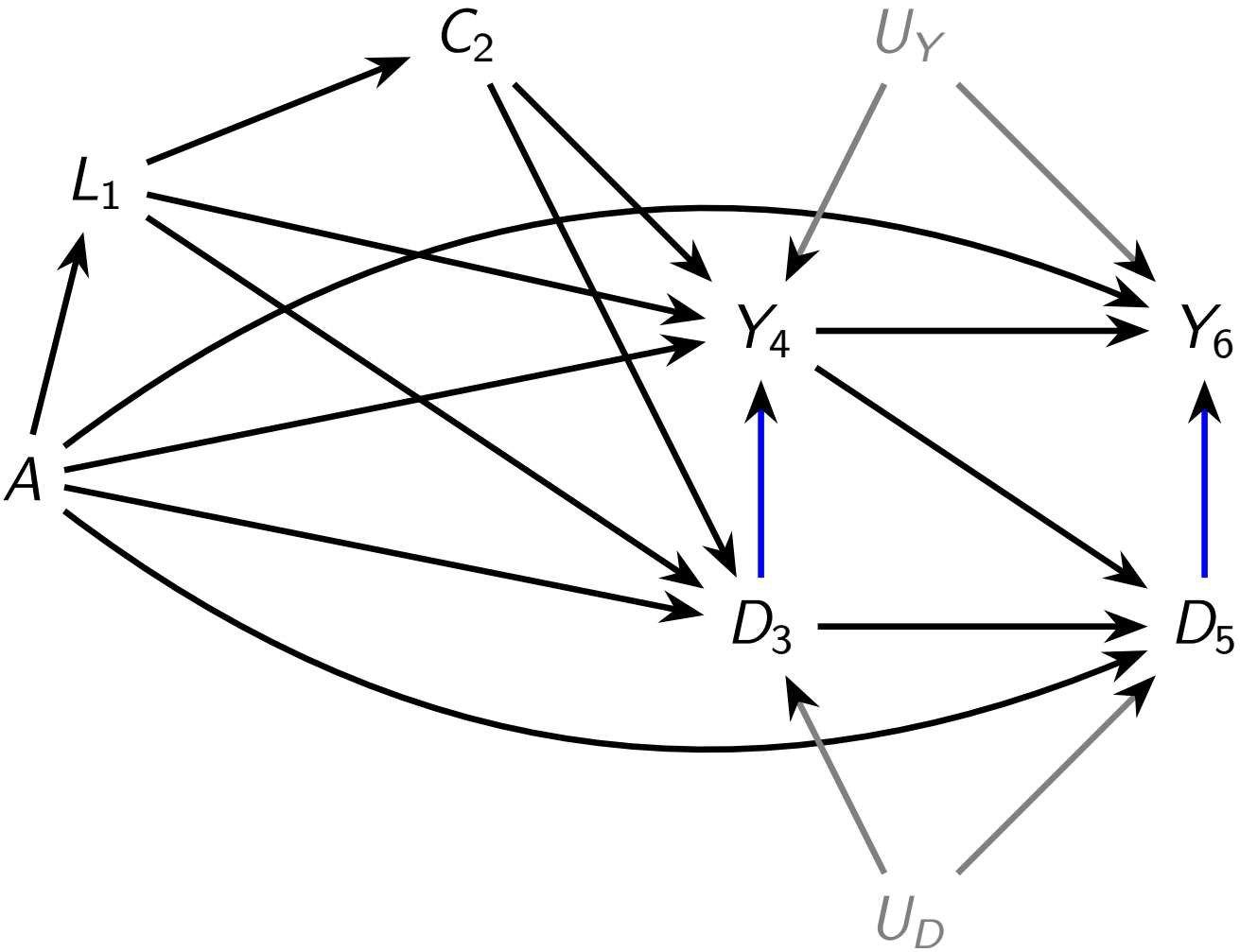
To avoid clutter, D nodes prior to C not depicted but all arguments still follow with these added as common causes of C and future Y.

Causal DAG depicting loss to follow-up



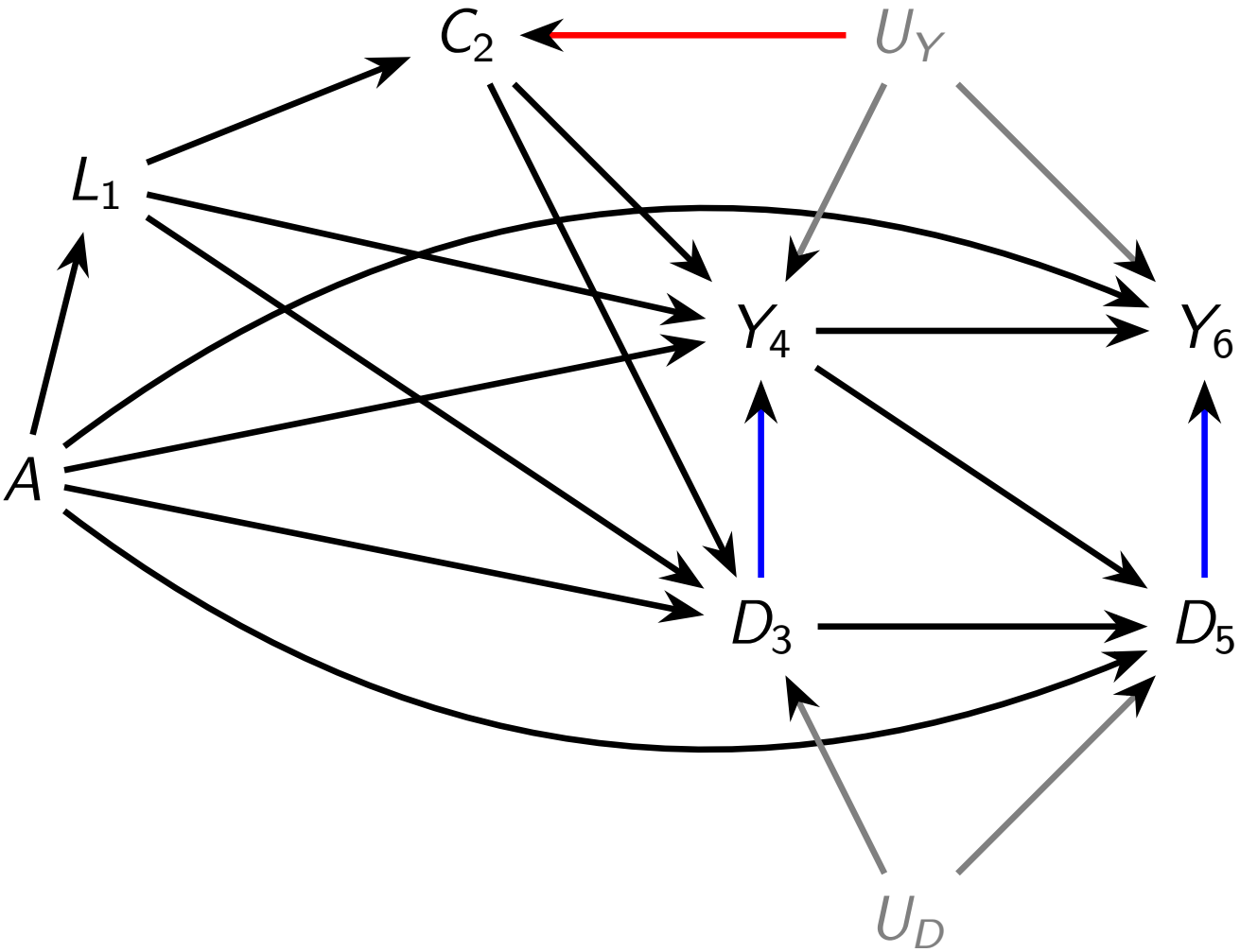
Can read exchangeability by "no unblocked backdoor paths" between V_k and future Y nodes conditional on "measured past".

Exchangeability for total effects



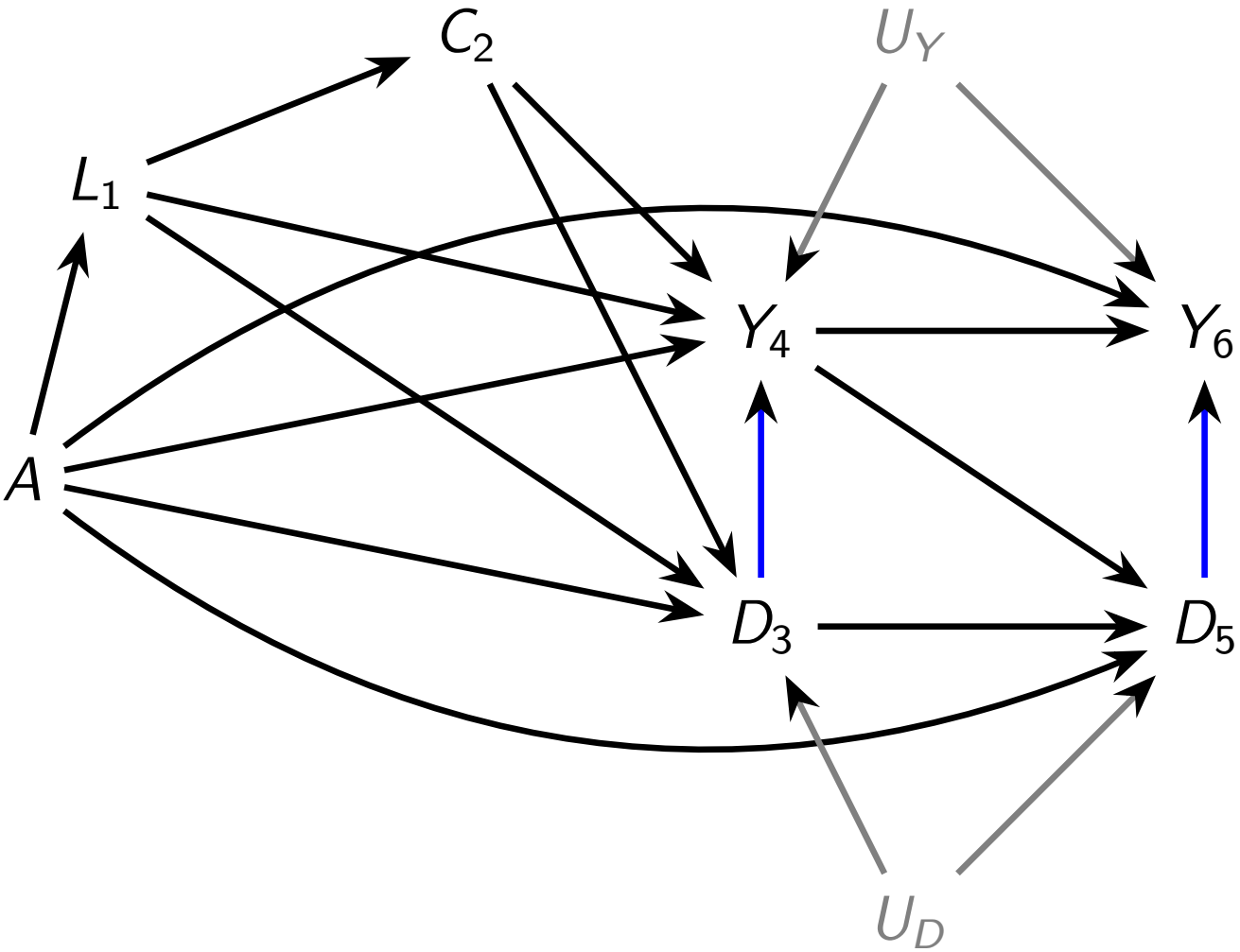
Need "no unblocked backdoor paths" between C_2 and future Y nodes conditional on "measured past"

Exchangeability for total effects



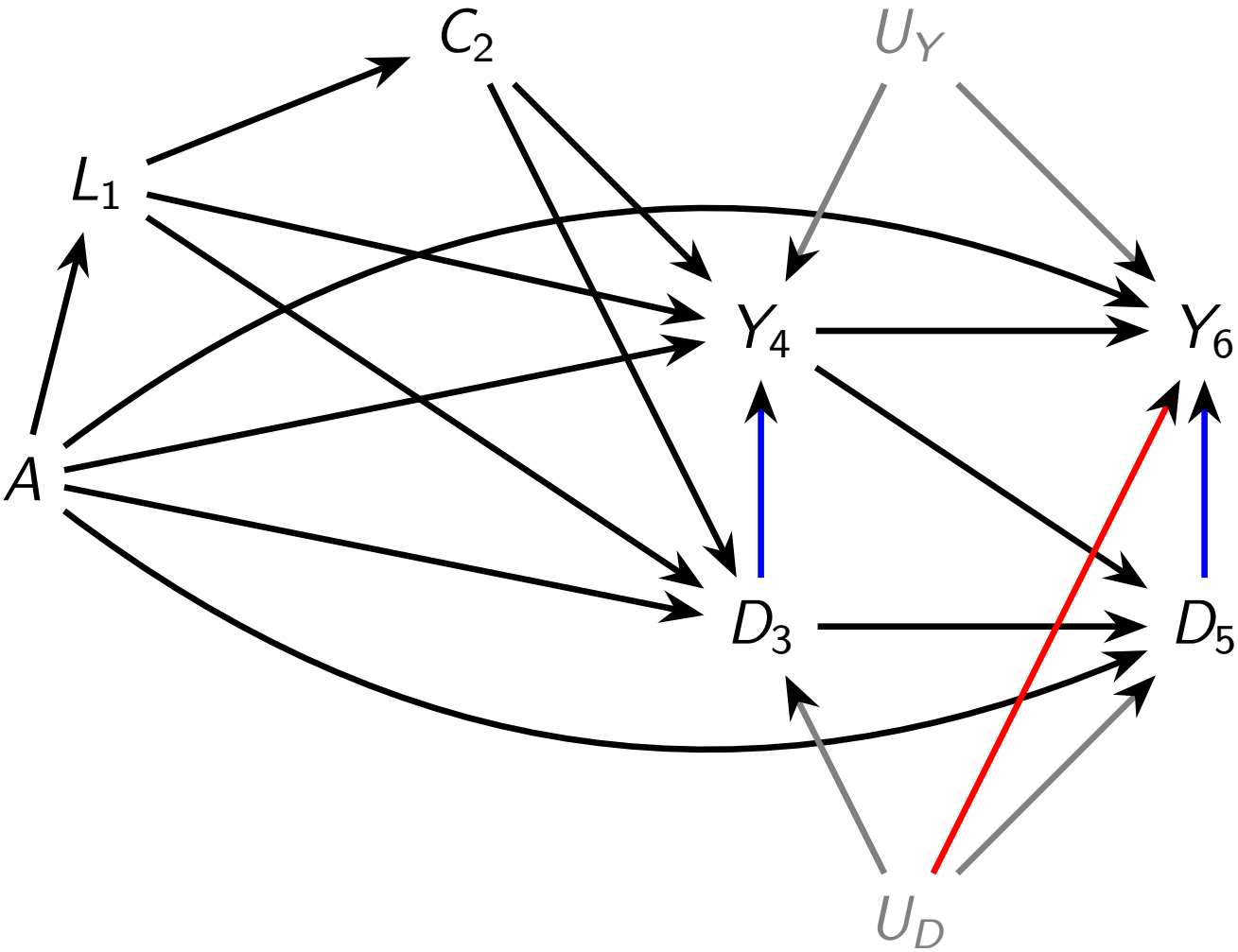
Need "no unblocked backdoor paths" between C_2 and future Y nodes conditional on "measured past"

Exchangeability for controlled direct effect



Need "no unblocked backdoor paths" between C_2, AND between each D node, and future Y nodes conditional on "measured past"

Exchangeability for controlled direct effect



Need "no unblocked backdoor paths" between C_2, AND between each D node, and future Y nodes conditional on "measured past"

Consistency and positivity for censoring

In order to identify total and controlled direct effects in censored data, we need more than just conditional exchangeability.

- Positivity condition
- Consistency condition

Like exchangeability these will be specific to the counterfactual target we are considering.

Positivity with respect to censoring

For any possible level of the "selection factor history" among **uncensored** and surviving individuals up to a particular time, it must be possible in the factual data to have **uncensored** individuals at the next time.

- This assumption ensures that our statistical parameter is well defined; that is, this parameter does not depend on conditional probabilities with empty denominators.

Consistency with respect to censoring

If an individual is uncensored up to k then their future counterfactual outcomes under intervention "eliminate censoring" up to k equal their factual ones.

- Consistency is generally not controversial when there is consensus on what a counterfactual outcome means (and what it doesn't mean).

Identifying function (the statistical parameter)

Under these conditions **for a definition of censoring**, we can write the risk of the event of interest by k under intervention $a, \overline{v} = 0$ $\Pr[Y_k^{a,\overline{v}=0} = 1]$ as function of only measured variables.

$$\sum_{\bar{l}_{k-1}} \sum_{j=0}^{k} \Pr[Y_j = 1 | \overline{L}_{j-1} = \bar{l}_{j-1}, \overline{V}_j = \overline{Y}_{j-1} = 0, A = a] \times$$

$$\prod_{s=0}^{j-1} \{\Pr[Y_{s-1} = 0 | \overline{L}_{s-2} = \bar{l}_{s-2}, \overline{V}_{s-1} = \overline{Y}_{s-2} = 0, A = a] \times$$

$$f(l_{s-1} | \bar{l}_{s-2}, \overline{V}_{s-1} = \overline{Y}_s = 0, a)\}$$

with competing events either components of L or V vectors depending on the nature of censoring events (the question). In turn, contrasts for different levels of a identify total or controlled direct effects.

Identifying function when D is a "selection factor"

Incorporating this more explicit notation, this is

$$\sum_{\bar{l}_{k-1}} \sum_{j=0}^{k} \Pr[Y_j = 1 | \bar{L}_{j-1} = \bar{l}_{j-1}, \overline{C}_j = \overline{D}_j = \overline{Y}_{j-1} = 0, A = a] \times$$

$$\prod_{s=0}^{j-1} \{\Pr[Y_{s-1} = 0 | \bar{L}_{s-2} = \bar{l}_{s-2}, \overline{C}_{s-1} = \overline{D}_{s-1} = \overline{Y}_{s-2} = 0, A = a] \times$$

$$\Pr[D_s = 0 | \bar{l}_{s-1}, \overline{C}_s = \overline{D}_{s-1} = \overline{Y}_{s-1} = 0, a] \times$$

$$f(l_{s-1} | \bar{l}_{s-2}, \overline{C}_{s-1} = \overline{D}_{s-1} = \overline{Y}_{s-1} = 0, a)\}$$

which is a function of conditional factual cause-specific hazards of the event of interest AND cause-specific hazards of the competing event. We get rid of the sum over $\bar{d}_{k-1}$ history by the determinism that defines the competing events data structure.

Identifying function when D is a "censoring event"

$$\sum_{\bar{l}_{k-1}} \sum_{j=0}^{k-1} \Pr[Y_j = 1 | \bar{L}_{j-1} = \bar{l}_{j-1}, \overline{C}_j = \overline{D}_j = \overline{Y}_{j-1} = 0, A = a] \times$$

$$\prod_{s=0}^{j-1} \{\Pr[Y_{s-1} = 0 | \bar{L}_{s-2} = \bar{l}_{s-2}, \overline{C}_{s-1} = \overline{D}_{s-1} = \overline{Y}_{s-2} = 0, A = a] \times$$

$$f(l_{s-1} | \bar{l}_{s-2}, \overline{C}_{s-1} = \overline{D}_{s-1} = \overline{Y}_{s-1} = 0, a)\}$$

This is only a function of the factual cause-specific hazards of the event of interest, not the any hazard of the competing event.

Statistics!

Now that we have a statistical parameter, we can discuss statistical methods. For example, our statistical parameter can be rewritten as a function of time-varying inverse probability of censoring weighted (IPCW) cause-specific hazards. This motivates weighted versions of familiar statistics from the survival analysis literature:

- IPCW version of the Aalen-Johansen estimator will converge to our statistical parameter for D a selection factor for censoring
- IPCW version of the Kaplan-Meier survival estimator will converge to our statistical parameter for D a censoring event.

Special cases

- When we can argue marginal exchangeability for censoring relative to the L's then we can justified unweighted versions of these popular statistics (but very unrealistic when censoring includes death!)
- When there is no loss to follow-up and competing events are not censoring events (target is total effect) then our complex identifying function reduces simply to $\Pr[Y_k = 1|A = a]$; that is, we are back to the idealized trial case.

Classic arguments for choosing between Aalen-Johansen and Kaplan-Meier

Widely cited pedagogic papers on competing risks analysis have given the argument against Kaplan-Meier in favor of Aalen-Johansen because Kaplan-Meier "overestimates" the risk of the event of interest but without any reference to a question!.

- Now you are able to really understand the choice here
- If you want a total effect then AJ is a valid approach
- If you want a controlled direct effect then KM is a valid approach
- Unweighted AJ justified only if marginal exchangeability for loss to follow-up is reasonable
- Unweighted KM is almost never reasonable because common causes of competing events and the event of interest will almost always be present.

Statistical assumptions

An additional assumption that these IPCW estimators make is a parametric model for the hazards of censoring at each time conditional on the past measured covariates:

- For the case where competing events not censoring this only requires a model for $Pr[C_{k+1} = 1 | \overline{L}_k, \overline{C}_k = \overline{D}_k = \overline{Y}_k = 0, A = a]$
- For the case where competing events are censoring events this additionally requires a model for $Pr[D_{k+1} = 1 | \overline{L}_k, \overline{C}_{k+1} = \overline{D}_k = \overline{Y}_k = 0, A = a]$

Estimates of these censoring hazards then used to construct person-time weights.

- Going to call this type of model assumption a *statistical assumption* in the sense that it isn't something typically made using subject matter arguments.

Other statistics

- Other statistics can be considered for the same statistical parameter that make different types of statistical assumptions.
- These come from the fact that the statistical parameters we showed are actually special cases of Robins's **g-formula** where the time-varying censoring events act like "time-varying treatments" and the time-varying selection factors act like "time-varying confounders".

Methods for estimating a g-formula sometimes broadly called "g-methods": include IP weighted methods, parametric g-formula (g-computation), doubly/multiply robust methods (e.g. TMLE).

Parametric g-formula

The parametric g-formula could be used to target either version of our statistical parameter. This requires many more statistical assumptions than IPCW including models for the joint distribution of the time-varying covariates over time.

- *gformula R package* has options for survival outcomes and treating competing events as censoring events or not.
- McGrath et al. gfoRmula: An R Package for Estimating the Effects of Sustained Treatment Strategies via the Parametric g-formula. *Patterns*. 2020.

Doubly/multiply robust methods

Doubly/multiply robust approaches can also be applied, rely on weaker statistical assumptions than the other methods:

- Any paper on these methods for survival outcomes with "censoring" in the data could be used to target a CDE by just treating competing events as censoring events. Example
  - Schnitzer et al. Modeling the impact of hepatitis C viral clearance on end-stage liver disease in an HIV co-infected cohort with targeted maximum likelihood estimation. *Biometrics*. 2014.
- Fewer papers that more explicitly target our statistical parameter where D is not conceptualized a censoring event but a "selection factor" (i.e. targeting total effects). One recent example
  - Diaz et al. Causal survival analysis under competing risks using longitudinal modified treatment policies. https://arxiv.org/abs/2202.03513. 2022.

Intro code resource

Over to Paloma!
`https://palolili23.github.io/2022_ser_competing_events/R/index.html`