Section 16

Lecture 5

My take on statistical science

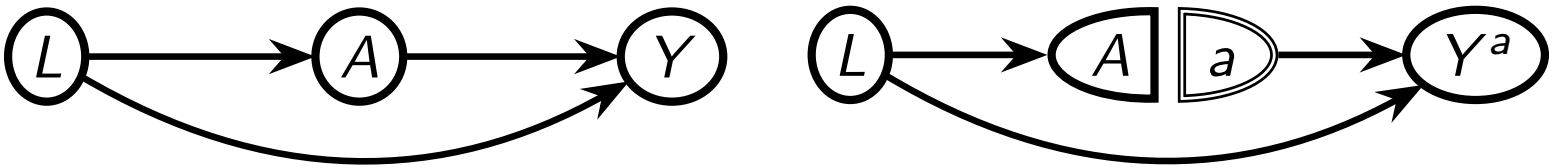
1. Start with the question.
   (Design your target trial)
2. Formalize the question in mathematical language.
   (Define your estimand)
3. Display the assumptions that are needed to identify your estimand.
   (Present your identifiability conditions)
4. Compute estimates of your estimands from your data.
   (Do your estimation)

$\Longrightarrow$ we **never** start the process by considering a regression model (linear, logistic, Cox model, ..., whatever).

Plan for today

- More on SWIGs
  - Examples
  - time-varying treatment
  - Clarifications
  - Proof of simple g-formula.
  - D-separation, g-formula and hidden variables.
  - Minimal labelling.
  - Read off independencies.
- Dynamic SWIGs.
- Next time: Estimation

SWIG in a conditional randomised experiment

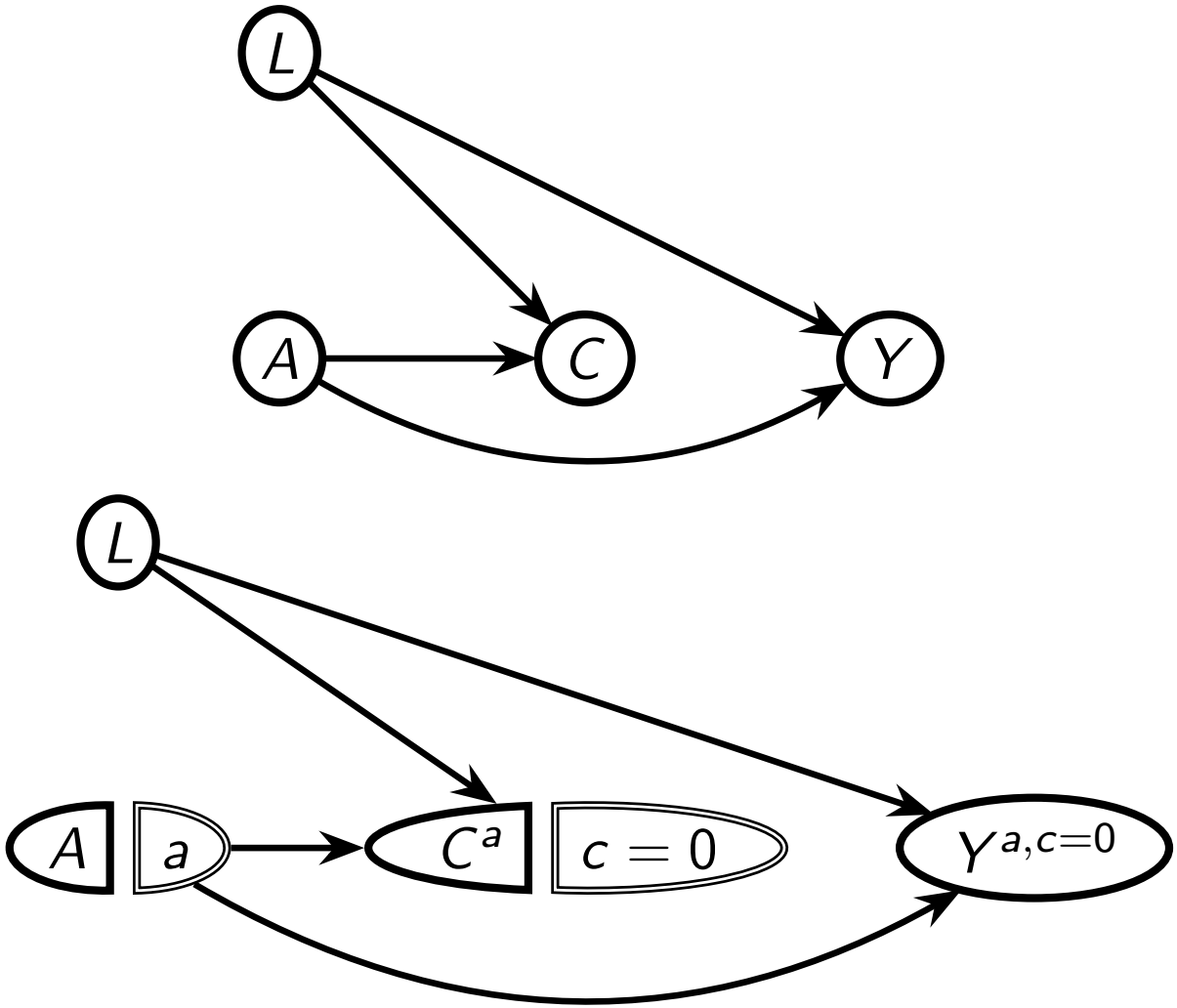


$$P(Y^a = y) = \sum_l P(Y^a = y \mid L = l)P(L = l) \text{ factorization}$$

$$= \sum_l P(Y = y \mid A = a, L = l)P(L = l). \text{ modularity}$$

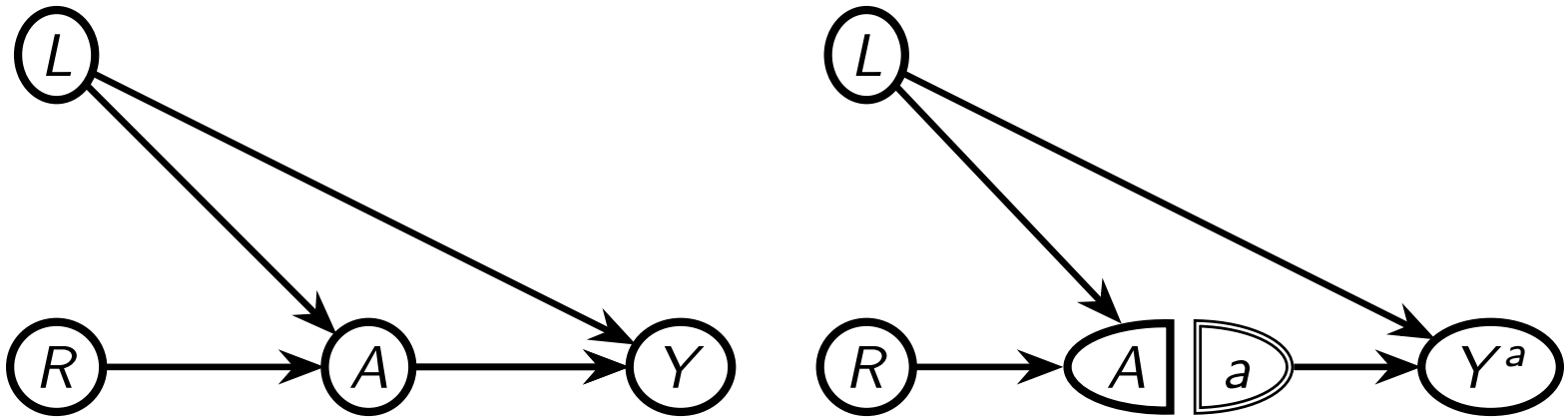
SWIG in an experiment with loss to follow-up (C)

A is treatment, C is censoring. The counterfactual outcome $Y^{a,c=0}$ is the outcome if we kept every individual uncensored ($c = 0$) under treatment a.



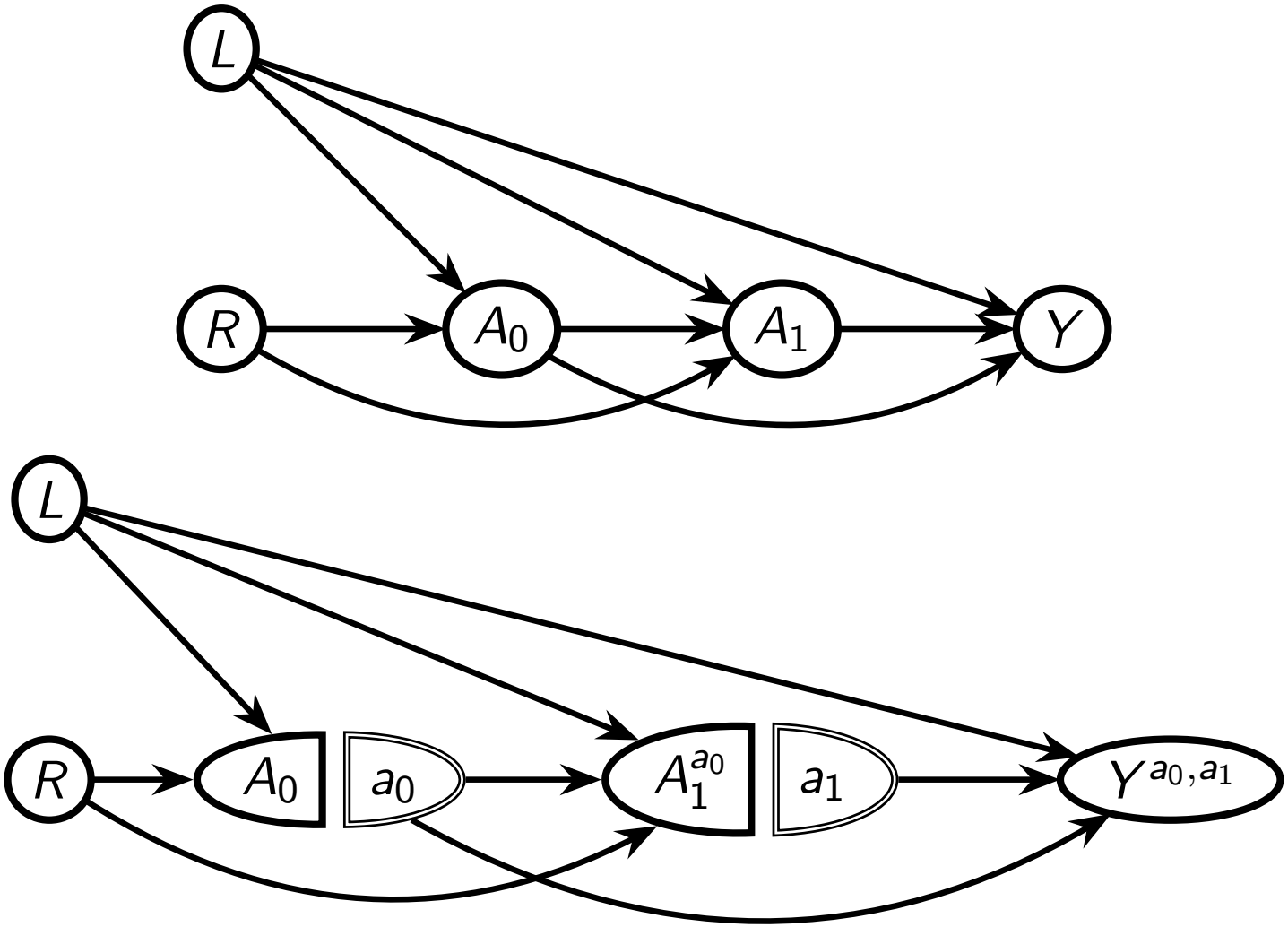
SWIG in an experiment with imperfect adherence

R is the strategy that was assigned, and A denotes taking treatment. Here, the counterfactual in the SWIG is the outcome had the patient taken treatment a. The lack of an arrow from R to Y^a encodes the assumption that randomisation only causes the outcome through the treatment A.



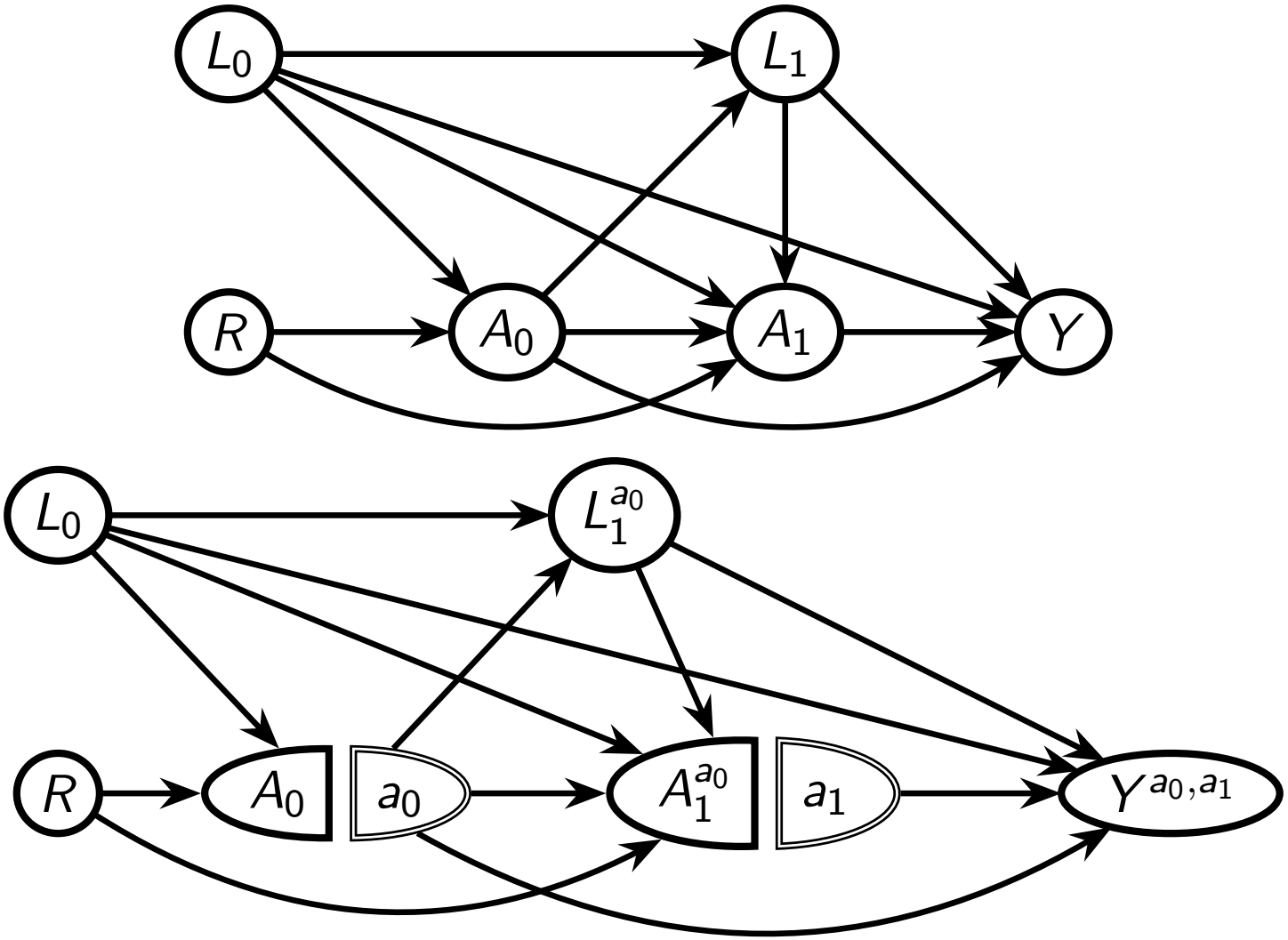
SWIG in an experiment with imperfect adherence

R is the strategy that was assigned, and A_k denotes taking treatment at time $k \in \{0, 1\}$.



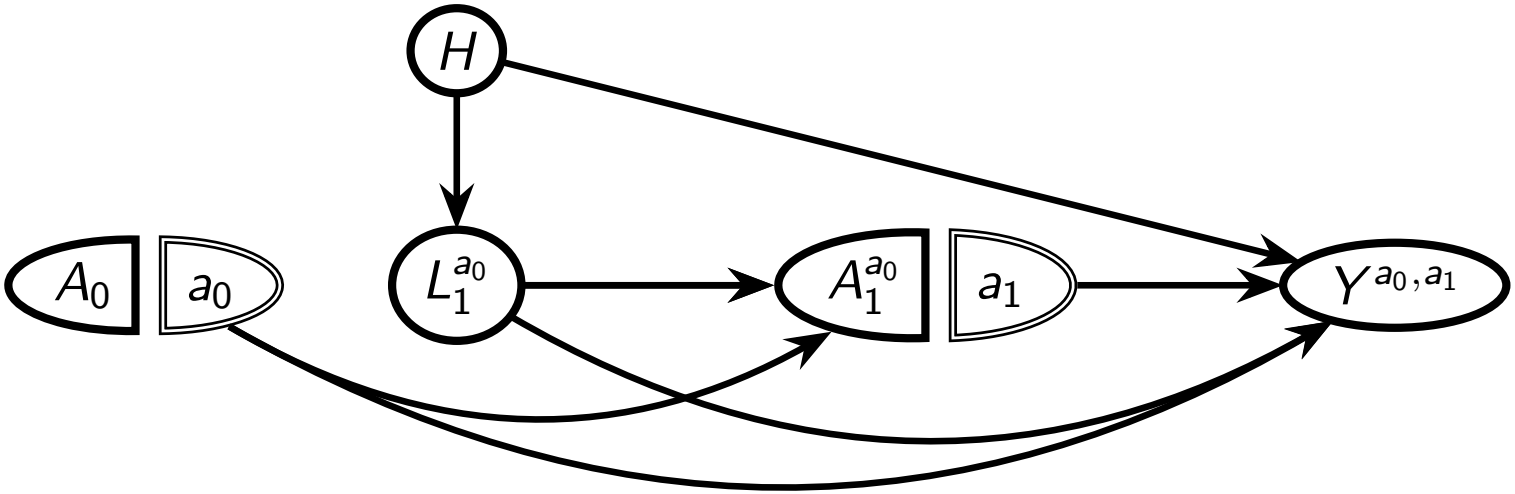
SWIG in an experiment with imperfect adherence

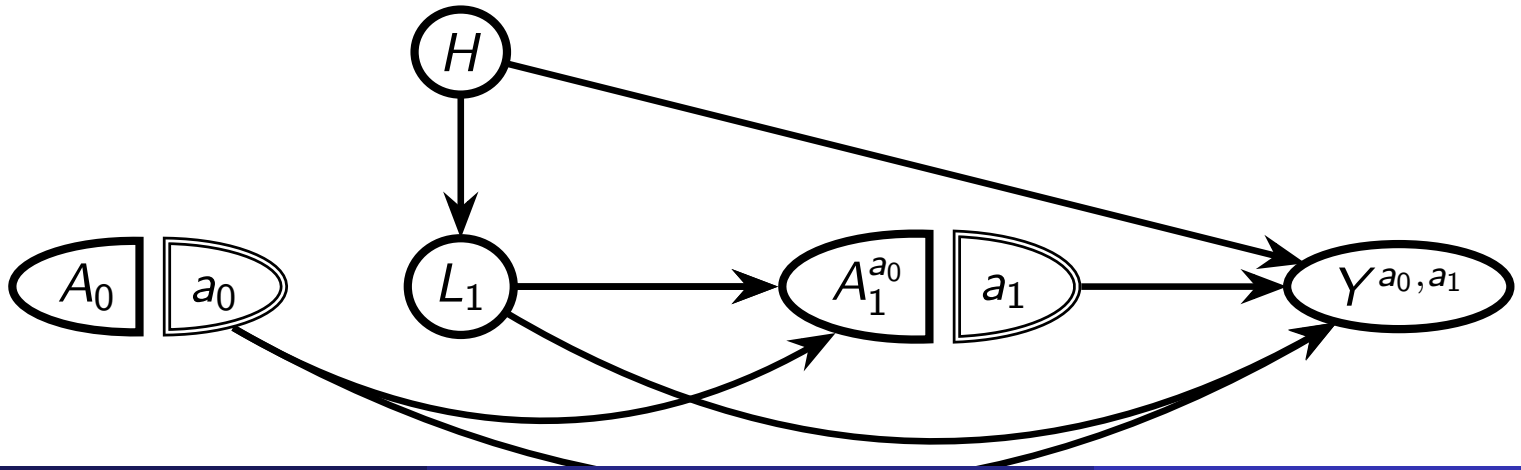
R is the strategy that was assigned, and A_k denotes taking treatment at time $k \in \{0, 1\}$.



SWIG and independencies

These graphs illustrate minimal labelling ($L_1^{a_0} = L_1$). The first graph is not minimally labelled, but encodes the same information as the second graph which is minimally labelled.





SWIG criterion for identification of effects

Consider the observed random variables $\overline{A}_K, \overline{L}_k, Y$.

Definition (marginal g-formula)

The g-formula for the *marginal* of $Y \equiv Y_K$ under treatment assignment $\overline{a} = \overline{a}_K = (a_0, \ldots, a_K)$ is defined as

$$b_{\overline{a}}(y) = \sum_{\overline{l}_K} p(y \mid \overline{l}_K, \overline{a}_K) \prod_{j=0}^{K} p(l_j \mid \overline{l}_{j-1}, \overline{a}_{j-1}),$$

where $\overline{l}_k = (l_0, \ldots, l_k)$, $k \leq K$, are instantiations of **observed** variables $\overline{L}_k = (L_0, \ldots, L_k)$, $k \leq K$.

We define variables indexed by subscript "-1", e.g. L_{-1}, to be empty.[29]

[29] Robins, "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect"; Richardson and Robins, "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality".

Note on the term "causal interpretation"

If it is

- A causal effect
- Equal to a counterfactual outcome of interest

Sufficient condition for identification

Theorem (Identification of static regimes)

Consider an intervention that sets $\bar{a} = \bar{a}_K = (a_0, \ldots, a_K)$. *Under positivity and consistency,*

$$P(Y^{\bar{a}} = y) = b_{\bar{a}}(y)$$

if for $k \in \{0, \ldots, K\}$

$$Y^{\bar{a}} \perp\!\!\!\perp I(A_k = a_k) \mid \bar{L}_k, \bar{A}_{k-1} = \bar{a}_{k-1}.$$

This theorem follows from Robins[30] and Richardson and Robins[31], and is closely related to the backdoor theorem of Pearl[32].
The theorem establishes when we can use the g-formula to identify causal effects.

---

[30] Robins, "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect".

[31] Richardson and Robins, "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality".

[32] Judea Pearl. "Causal diagrams for empirical research". In: *Biometrika* 82.4 (1995), pp. 669–688.

Proof in a simple case

Consider the case with two treatments (A_0, A_1) and a binary outcome $Y \in \{0, 1\}$. Suppose that $Y^{a_0,a_1} \perp\!\!\!\perp A_0$ and $Y^{a_0,a_1} \perp\!\!\!\perp A_1 \mid L_1, A_0 = a_0$

Proof.

$$
\begin{aligned}
\mathbb{E}(Y^{a_0,a_1}) &= \mathbb{E}(Y^{a_0,a_1} \mid A_0 = a_0) \text{ exchangeability} \\
&= \sum_{l_1} \mathbb{E}(Y^{a_0,a_1} \mid L_1 = l_1, A_0 = a_0) p(l_1 \mid a_0) \text{ LTOT} \\
&= \sum_{l_1} \mathbb{E}(Y^{a_0,a_1} \mid A_1 = a_1, L_1 = l_1, A_0 = a_0) p(l_1 \mid a_0) \text{ exchangeability} \\
&= \sum_{l_1} \mathbb{E}(Y \mid A_1 = a_1, L_1 = l_1, A_0 = a_0) p(l_1 \mid a_0) \text{ consistency, positivity}
\end{aligned}
$$

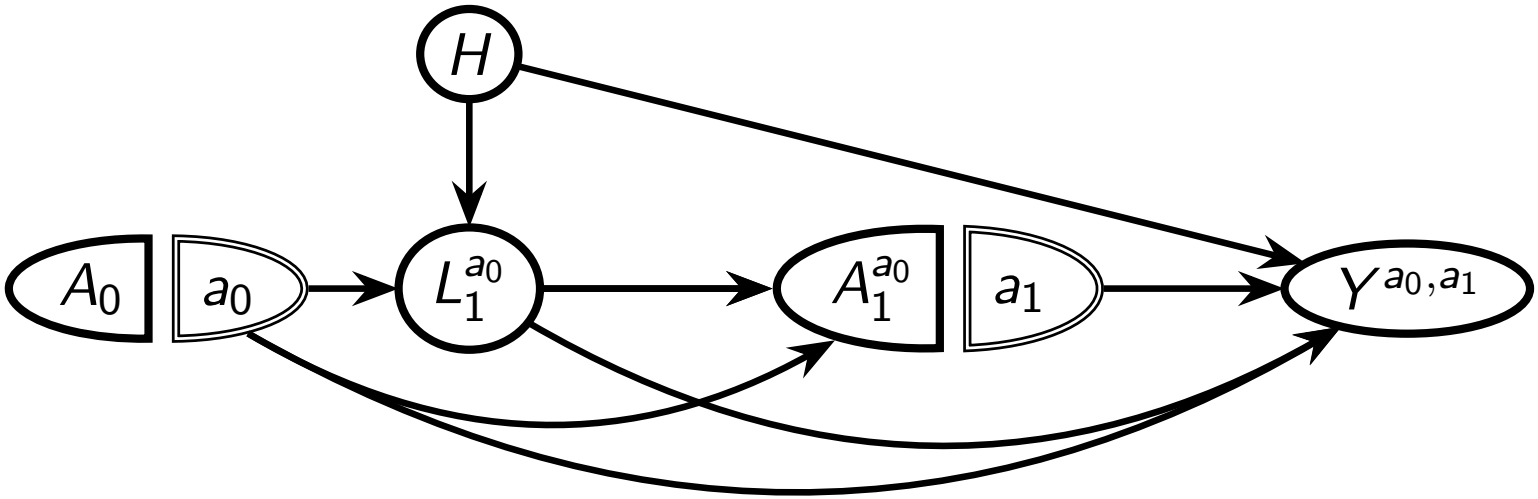
□

Comments to the g-formula

- The independence condition in the identification theorem cannot be read directly off of a SWIG. However, on the next slide we see how the identification condition is implied by an independence in the SWIG.
- Importantly, the g-formula allows identification in the presence of unmeasured variables.

Reading off independencies in SWIGs

Let H be a hidden (unmeasured) variable



We can read off $Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid L_1^{a_0}, A_0$.
However, what we needed for using the g-formula is the independence $Y^{a_0,a_1} \perp\!\!\!\perp A_1 \mid L_1, A_0 = a_0$.
Use consistency: $A_1^{a_0} \mid L_1^{a_0}, A_0 = a_0$ is equal to $A_1 \mid L_1, A_0 = a_0$, i.e.,
$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid L_1^{a_0}, A_0 \implies Y^{a_0,a_1} \perp\!\!\!\perp A_1 \mid L_1, A_0 = a_0$.

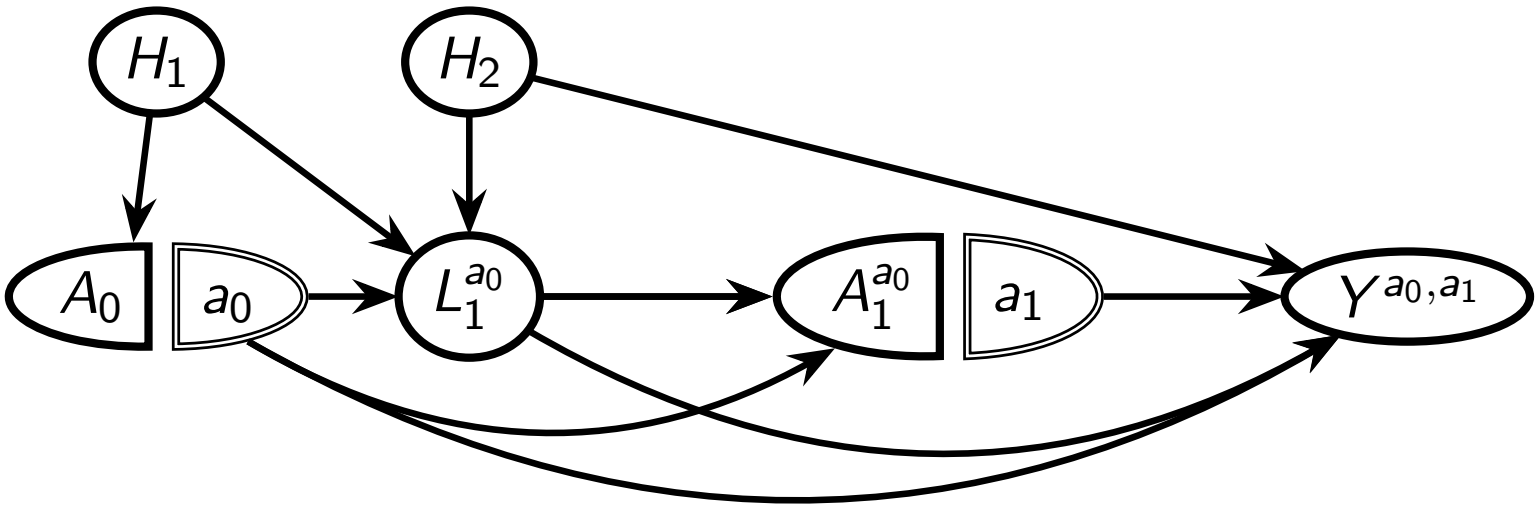
Using the identification theorem

Thus, we can identify the expected counterfactual outcome under the intervention that sets $A_0 = a_0$ and $A_1 = a_1$ in the graph in Slide 167 as

$$\mathbb{E}(Y^{a_0,a_1}) = \sum_{l_1} \mathbb{E}(Y \mid A_1 = a_1, L_1 = l_1, A_0 = a_0) P(L_1 = l_1 \mid A_0 = a_0).$$

Note that we have identified the counterfactual as a function of only the observed variables in the graph, even if there is a hidden variable H in the graph.

Additional SWIG



What is the g-formula? Compare to Figure 167. Indeed, the g-formula is just a function of observed data distributions, but here we have no guarantee that it does identify the causal estimand because the identification conditions are violated.

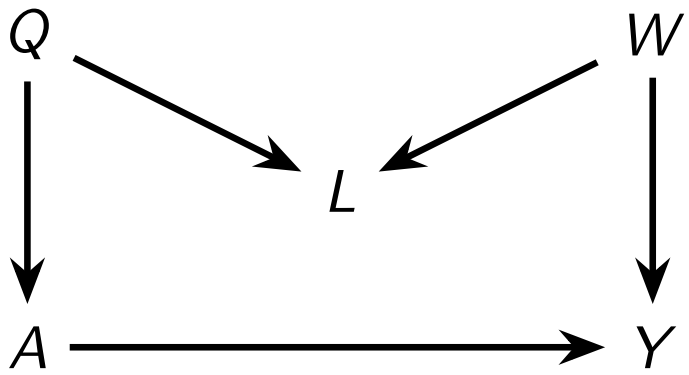
Some insights

- We have studied identification from an "all or nothing" perspective.
  - We will later look at sensitivity analyses and bounds.
- The identification assumptions we have studied are non-parametric (PS: I consider this to be a feature, not a bug). We have not considered other assumptions that also can be used to justify identification, for example
  - monotone effects.
  - no effect modification.
- We have **not learned** the graphical structure. On the other hand, we have learned what we can infer from a given graphical structure; heuristically, we encode what we know and believe in the graph, and then we deduce what we can learn from this knowledge and assupmtions.
  - Learning the graphical structure itself from data is a very ambitious task.
  - In principle, the causal structure could be learned by doing a large amount of experiments (I am not discussing this in more detail here).

Hidden variables

- Importantly, the g-formula allows identification in the presence of hidden variables.

Berkson bias



- A Drink a glass of red wine a day.
- Y Nausea
- L Aspirin
- Q Family history of cardiovascular disease
- W Frequency of headache

Q: We measure Aspirin. Should we adjust for Aspirin in the analysis? Draw the SWIG...

Section 17

Dynamic regimes

Dynamic regimes

Definition (Dynamic regime)

A dynamic regime $g = (g_0, \ldots, g_k)$, where $g_k : (\overline{A}_{k-1}, \overline{L}_k) \mapsto A_k$, is a policy that assigns treatment (possibly at multiple time points) based on the measured history $(\overline{A}_{k-1}, \overline{L}_k)$.

We will restrict ourselves to settings where

$$g_k : (\overline{L}_k) \mapsto A_k$$

.

Dynamic regime SWIGs

Definition (d-SWIG from Robins and Richardson)

Given a template $\mathcal{G}(a)$ and a dynamic regime g for $\bar{a}$, the d-SWIG $\mathcal{G}(g)$ is defined by applying the following transformation:

- Replace each fixed node a_j with a random node A_j^{g+} that inherits children from a_j. Include dashed directed edges from every variable that is an input to the function g_i that determines the variable A_i^{g+}.
- Each random node V_i that is a descendant of at least one variable A_i^{g+} is relabeled as V_i^g.

Time-varying exposures (treatments) are frequent

Examples:

- Smoking status, which depends on other events in life.
- A therapeutic drug, for which the dose is adjusted according to the response over time (patients take the drug every day, every week etc)
- Cancer screening, which e.g. depends on previous diagnostic tests.
- Surgical interventions (e.g. transplants) are given at a certain time after the diagnosis.
- Expression of genes.

Running example: HIV

Consider a 5-year follow-up study of individuals infected with the human immunodeficiency virus (HIV)[33].

- A_k takes value 1 if the individual receives antiretroviral therapy in month k, and 0 otherwise. Define $A_{-1} = 0$.
- Suppose Y measures health status at 5 years of follow-up.
- So far we have considered *deterministic* treatment rules, for example "always treat", where the outcome of interest is $Y^{a=1}$ vs "never treat", where the outcome of interest is $Y^{a=0}$.
  When $\overline{A} \equiv \overline{A}_K$, we can define 2^K such static regimes...
- However, often we want to make *dynamic* treatment decisions.
- Let $L_k \in \{0, 1\}$ be an indicator of low CD4 cell count measured at month k.
- Depending on the value of L_k, we may argue that it is good or bad to start treatment at time k.

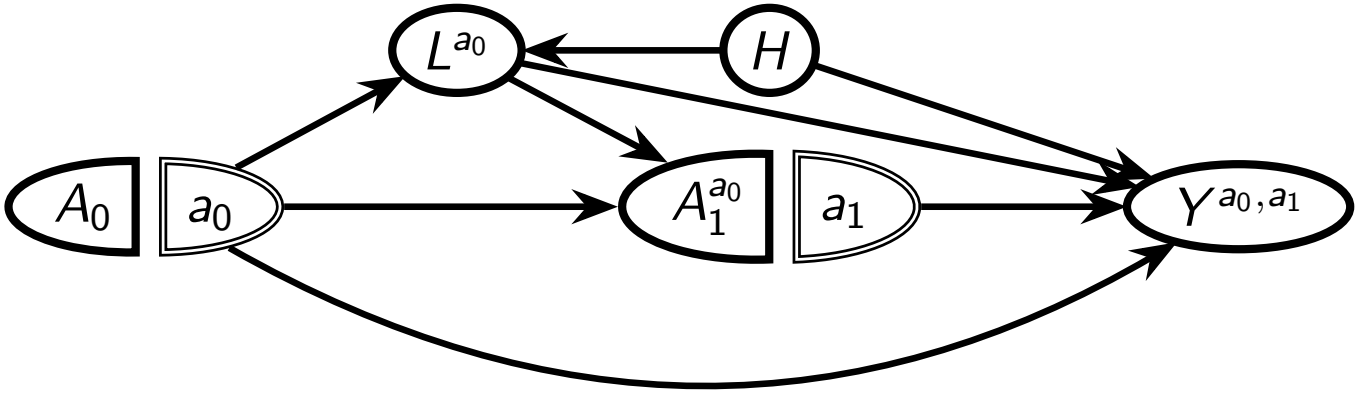
[33] Hernan and Robins, *Causal inference: What if?*

Example of Dynamic Regime

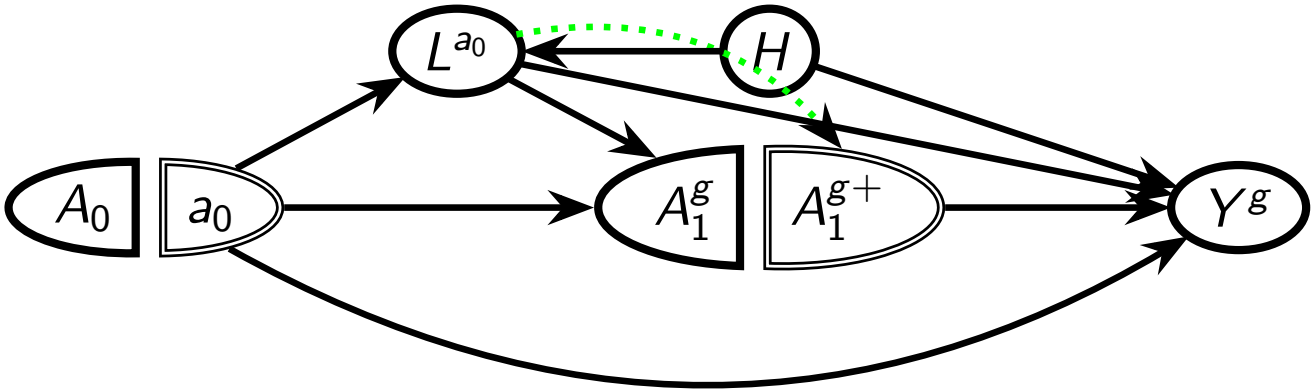
A simple example of a dynamic regime g for setting with two treatments is

- $A_0^{g+} = a_0$.
- $A_1^{g+} = L_1^{a_0}$
  In the HIV example this would mean that you are treated at time 1 if the CD4 cell count is low at that time.

Static vs dynamic



$Y^{a_0,a_1} \perp\!\!\!\perp A_0$ and $Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid L_0^{a_0}, A_0$.



$Y^{g} \perp\!\!\!\perp A_0$ and $Y^{g} \perp\!\!\!\perp A_1^{a_0} \mid L_0^{a_0}, A_0$.
Consistency gives: $Y^{g} \perp\!\!\!\perp A_0$ and $Y^{g} \perp\!\!\!\perp A_1 \mid L_0, A_0 = a_0$.

Identification results for dynamic regimes

- We can use the same identification conditions (independencies in Slide 164) as for static regimes, only if A_k^{g+} **is not** a function of A_j^{g+} for $j < k$; that is, A_k^{g+} cannot be written a function of only $\overline{L}_k$. However, we need to use the extended g-formula as the identification formula (as defined in Slide 186).
- if A_k^{g+} **is** a function of A_j^{g+} for any $j < k$, we need slightly stronger conditions (we are not presenting them now). This is e.g. the case in the graph in Slide 184 (due to the red arrow).