Methods for causal inference from observational data (I)

James Robins, Miguel A. Hernán

Departments of Epidemiology and Biostatistics

Harvard School of Public Health

An intuitive definition of cause

- Jim didn’t take the pill on Sept 1, 2001
  - Five days later, he was alive

- Had Jim taken the pill on Sept 1, 2001 (all others things being equal)
  - Five days later, he would have been alive

- Did the pill cause Jim’s survival?

Notation for actual data

- Y=1 if patient died, 0 otherwise
  - Y_i=1, Y_j=0
- A=1 if patient treated, 0 otherwise
  - A_i=1, A_j=0

ID	A	Y
Ian	1	1
Jim	0	0

Notation for ideal data

- $Y_{a=0}$=1 if patient would have died, had he *not* taken the pill
  - $Y_{i,\ a=0}=1,\ Y_{j,\ a=0}=0$
- $Y_{a=1}$=1 if patient would have died, had he taken the pill
  - $Y_{i,\ a=1}=1,\ Y_{j,\ a=1}=0$

ID	A	Y	$Y_{a=0}$	$Y_{a=1}$
Ian	1	1	1	1
Jim	0	0	0	0

(Individual) Causal effect

- For Ian:
    - Pill has a causal effect if $Y_{i,\ a=0} \neq Y_{i,\ a=1}$
- For Jim:
    - Pill has a causal effect if $Y_{j,\ a=0} \neq Y_{j,\ a=1}$

- Unfortunately, individual causal effects cannot be determined because…

Available data on Ian and Jim

ID	A	Y	$Y_{a=0}$	$Y_{a=1}$
Ian	1	1	?	1
Jim	0	0	0	?

- *If untreated, A=0,* $Y = Y_{a=0.}$ and $Y_{a=1}$ is missing
- *If treated, A=1,* $Y = Y_{a=1}$ and $Y_{a=0}$ is missing
- *We refer to* $Y_{a=0}$ and $Y_{a=1}$ as counterfactual outcomes.
- Counterfactual and actual outcomes are linked via

$$Y_A = Y$$

For a binary (0,1) variable, E[Y]=pr(Y=1)
The average of Y in the population is the proportion of 1's.

	Your table		God's table		
ID	*A*	*Y*	*A*	$Y_{a=0}$	$Y_{a=1}$
1	0	0	0	0	0
2	0	1	0	1	1
3	0	0	0	0	0
4	0	0	0	0	0
5	1	1	1	1	1
6	1	1	1	1	1
7	1	0	1	0	0
8	1	0	1	0	0
$\Pr(Y=1 \mid A=1) - \Pr(Y=1 \mid A=0) =$ $2/4 - 1/4 = 1/4$			$\Pr(Y_{a=1}=1) - \Pr(Y_{a=0}=1) =$ $E(Y_{a=0} - Y_{a=1}) = 0$		

Association ≠ Causation: Confounding for the effect of A on Y

Association vs Causation

- Association (identifiable)

$$\alpha = E(Y|A=1) - E(Y|A=0) = E(Y_{a=1}|A=1) - E(Y_{a=0}|A=0)$$

crude risk difference

- Causation (not identifiable)

$$\theta = E(Y_{a=1} - Y_{a=0}) = E(Y_{a=1}) - E(Y_{a=0})$$

causal risk difference=
average causal effect

proportion diseased if all treated -
proportion diseased if all untreated

Average of a difference (sum) is the
difference (sum) of the averages.

Association vs Causation

- Under randomization

Treatment is independent of any fixed pretreatment variable under randomization. Counterfactuals, like genes, are existing pretreatment variables that can only be observed ater .
2x2 table of A versus either counterfactual has OR=RR=1 RD=0

$$(Y_{a=1}, Y_{a=0}) \amalg A$$

$$\alpha = \theta$$

Causation = Association

Treatment is not independent of Y if the null is false because Y is a function of A.
2x2 table of A vs Y has OR and RR differ from 1 and RD different from 0

- However, individual effect $Y_{a=1}$ - $Y_{a=0}$ is still not identified

Statistical Notation: and concepts

$$pr\left(A = a \middle| Y = y\right) = f_{A|Y}\left(a \middle| y\right) = f\left(a \middle| y\right).$$

Note $pr\left(A = 6 \middle| Y = 4\right) = f_{A|Y}\left(6 \middle| 4\right)$ but cannot write $f\left(6 \middle| 4\right)$ since we do not know which variable is meant.

A joint probability distribution for (A, Y) where $a \in \{1, 2, 3\}$ and $y \in \{0, 1, 2, 3\}$.

A/Y	0	1	2	3
1	.05	.06	.07	.08
2	.19	.11	.01	.13
3	.15	.01	.10	.04

1.What is $pr\left(A=1, Y=0\right)$? .05

1. Is this a probability distribution.?

Yes because

$\sum_{a,y} pr\left(A=a, Y=y\right)=$

$\sum_{y=0}^{y=4}\left\{\sum_{a=1}^{a=3} pr\left(A=a, Y=y\right)\right\}=$

$\sum_{a,y} f\left(a, y\right)=1$ ie the sum of all entries in the table is one.

2. $pr\left(A=2\right)=$

$\sum_{y=0}^{y=4} pr\left(A=2,Y=y\right)=$

$.19+.11+.01+.13..$

3.$More$ generally

$pr\left(A=a\right)=f\left(a\right)=\sum_{y=0}^{y=4} f\left(a,y\right)$

4. $E(A) =$

$\sum_a a pr(A = a) =$

$\sum_a a f(a) =$

$\sum_a a \left\{ \sum_{y=0}^{y=4} f(a, y) \right\} =$

$\sum_a \sum_{y=0}^{y=4} a f(a, y) =$

5. $f\left(a|y\right) =$

$pr\left(A = a|Y = y\right) =$

$pr\left(A = a, Y = y\right)/pr\left(Y = y\right) =$

$pr\left(A = a, Y = y\right)/\sum_a pr\left(A = a, Y = y\right) =$

$f\left(a, y\right)/\sum_a f\left(a, y\right)..$

$pr\left(A = 2|Y = 3\right) = .13/\left\{.08 + .13 + .04\right\}$

$f\left(a, y\right) =$

$f\left(a|y\right) f\left(y\right) =$

$f\left(y|a\right) f\left(a\right).$

$Check$ for $a = 2$ and $y = 3$.

If three variables (A, Y, L), consider all possible orderings

$$L \rightarrow Y \rightarrow A, Y \rightarrow L \rightarrow A, L \rightarrow A \rightarrow Y,$$

$$A \rightarrow L \rightarrow Y, A \rightarrow Y \rightarrow L, Y \rightarrow A \rightarrow L,$$

$$f(a, y, l) =$$

$$f(a|y, l) f(y|l) f(l) =$$

$$f(a|y, l) f(l|y) f(y) =$$

$$f(y|a, l) f(a|l) f(l) =$$

$$f(y|a, l) f(l|a) f(a) =$$

$$f(l|y, a) f(y|a) f(a) =$$

$$f(l|y, a) f(a|y) f(y).$$

Representation of conditional independence of dichotomous variables A and Y given L

- $A \perp\!\!\!\perp Y|L=l$ for all l, $A \perp\!\!\!\perp Y|L$ $OR_{AY|L}=RR_{AY|L}=RR_{YA|L}=1$, $RD_{AY|L}=RD_{YA|L}=0$
- $\Pr[A=1|Y=1,L]=\Pr[A=1|Y=0,L]$, $\Pr[Y=1|A=1,L]=\Pr[Y=1|A=0,L]$,
- $E[Y|A=1,L]=E[Y|A=0,L]$, $E[Y|A=1,L]=E[Y,L]$
- $E[Y|A=1,L]=E[Y|A=0,L]$, $E[Y|A=1]=E[Y]$
- $E[A|Y=1,L]=E[A|Y=0,L]$, $E[A|Y=1,L]=E[A|L]$
- $f(a|y,l)=f(a|l)$, $f(y|a,l)=f(y|l)$ forall l

Representation of independence of dichotomous variables A and Y

- $A \perp\!\!\!\perp Y$
- $OR_{AY}=RR_{AY}=RR_{YA}=1$,
  $RD_{AY}=RD_{YA}=0$
- $Pr[A=1|Y=1]=Pr[A=1|Y=0]$,
  $Pr[Y=1|A=1]=Pr[Y=1|A=0]$,
- $E[Y|A=1]=E[Y|A=0]$,
  $E[Y|A=1]=E[Y]$
- $E[Y|A=1]=E[Y|A=0]$,
  $E[Y|A=1]=E[Y]$
- $E[A|Y=1]=E[A|Y=0]$,
  $E[A|Y=1]=E[A]$
- $f(a|y)=f(a)$, $f(y|a)=f(y)$,

No unmeasured confounders

- Suppose

$$(Y_{a=1}, Y_{a=0}) \not\perp\!\!\!\perp A \quad (\text{so} \quad \alpha \neq \theta \quad)$$

- Suppose however,

$$(Y_{a=1}, Y_{a=0}) \perp\!\!\!\perp A \mid L$$

Means independence within each level of L. False if dependent at even one level

- Then we say that there are no unmeasured confounders for the effect of A on Y
  - L-Biased coin randomization
  - Observational studies where L are all important confounders (though in practice we will need sensitivity analysis)

G-formula

- Suppose $(Y_{a=1}, Y_{a=0}) \amalg A \mid L$, then

$$E(Y_{a=1}) = \sum_l E(Y_{a=1} \mid L = l) f_L(l)$$

law of total probability.
f(l)=pr(L=l)
Do not need sub L since l

average ht is average ht in men times the proportion of men plus av ht of women times proportion women

$$= \sum_l E(Y_{a=1} \mid A = 1, L = l) f_L(l)$$

By independence

$$= \sum_l E(Y \mid A = 1, L = l) f_L(l)$$

By definition of Y in terms of counterfactuals

G-formula is the direct standardization of

$$E(Y \mid A = 1, L)$$

Standardized Risk Difference

$$\theta \quad = \quad E(Y_{a=1} - Y_{a=0}) \quad =$$

$$= \quad \Sigma_l \{E(Y_{a=1} | A=1, l) - E(Y_{a=0} | A=0, l)\} f_L(l)$$

risk difference in statum L=l

weights is dist of L in total population

When is L a non confounder?

We assume you can get effect if L is measured so we need 1. below. When can we still get causal effect without data on L. If Conditon 2 holds
Mathematically when is the standardized RD equal to the crude RD. Answer: When 2 holds.

Suppose

1. $(Y_{a=1}, Y_{a=0}) \coprod A \mid L$
2. $L \coprod Y \mid A$ *or* $L \coprod A$

- Then

$$E\,(Y_{a=1}) = E(\,Y \mid A = 1)$$

Because

$$\textstyle\sum_l \{E(Y|A=1,l)\ f_L\,(l)\} = \sum_l \{E(Y|A=1,l)\ f_L\,(l \mid A=1)\}$$

=E(Y|A=1)

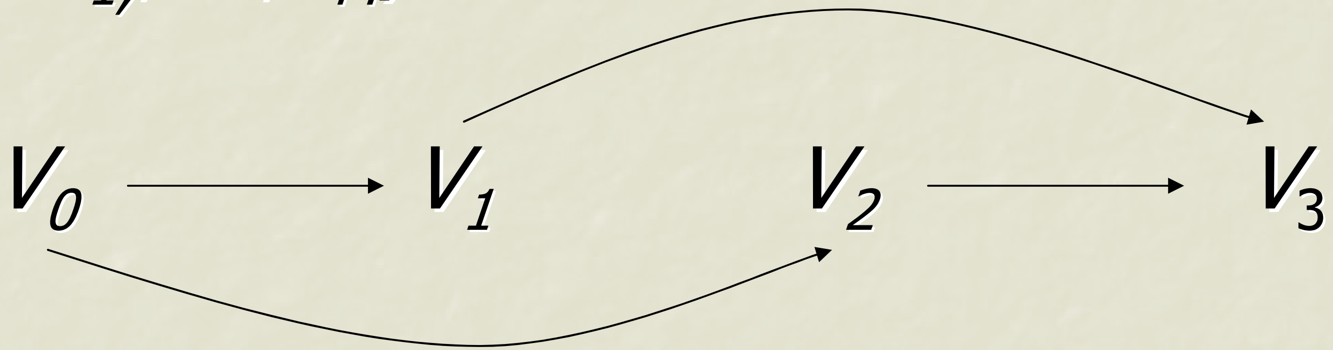
- If (1) and (2) hold, then L is a non-confounder and

$$(Y_{a=1}, Y_{a=0}) \coprod A \quad \textit{and} \quad E(Y_{a=1} - Y_{a=0}) = E(\,Y \mid A = 1) - E(\,Y \mid A = 0)$$

Conditonal on L
randomization plus 2. imply
unconditional randomization
(no confounding)

DAGS

- $(V_{0,}, V_{1,}, \ldots, V_M)$

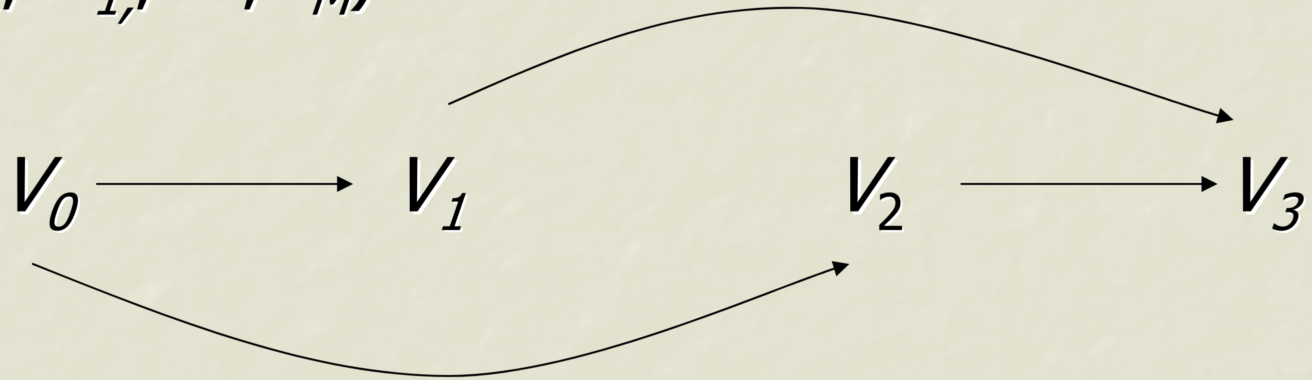


- Dags:
  - Nodes are random variables
  - Edges are directed arrows
  - No directed cycles

Think temporal ordering left to right. Arrows then must go from past to future.

DAGS

- $(V_0, V_1, \ldots, V_M)$



- Statistical Dag:

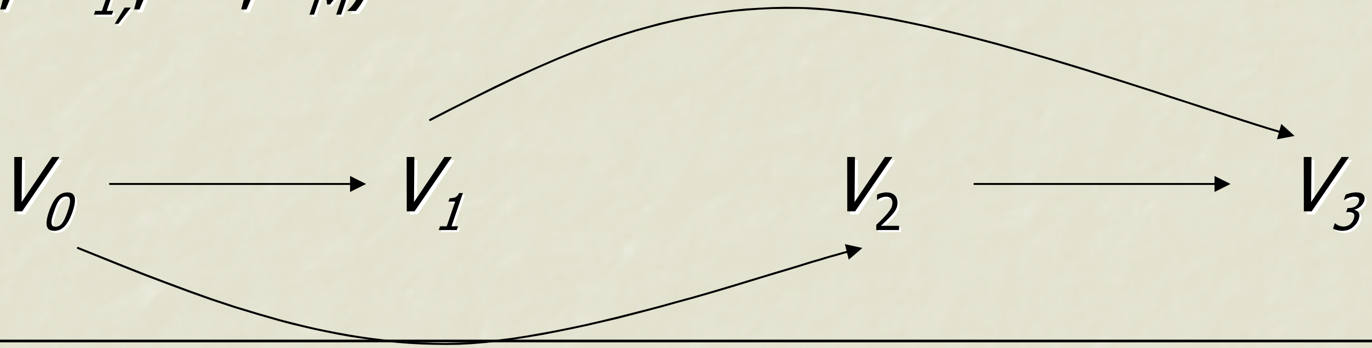
$$f(V) = \prod_{m=1}^{M} f(V_m | pa_m)$$ pa_m are the parents of V_m

- Example

$$f(V) = f(V_3 | V_1, V_2)\, f(V_2 | V_0)\, f(V_1 | V_0)\, f(V_0)$$

DAGS

- $(V_0, V_1, \ldots, V_M)$



statistical dag: equivalent definitions: Each variable is independent of all variables in its past given its parents
Each variable is independent of its nondescendants given its parents.
In this DAG V2 is independent of V1 given (within each level of) V0.

- Example

$$f(V) = f(V_3|V_1, V_2)\, f(V_2|V_0)\, f(V_1|V_0)\, f(V_0)$$

- Complete Dag

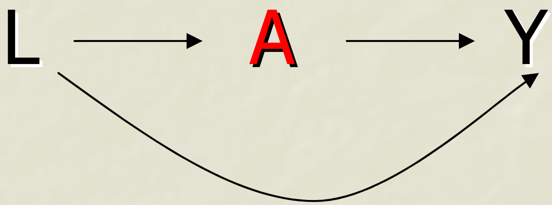
V2 is independent of V1 ie OR=1
with each level of V0 if the the probabilty V2 takes any value is not predicted by V1

Given V0, V1 is not an independent predictor of (risk factor for) V2

$$f(V) = f(V_3|V_0, V_1, V_2)\, f(V_2|V_0, V_1)\, f(V_1|V_0)\, f(V_0)$$

G-Formula and DAGS

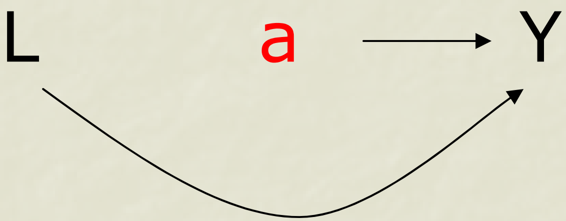
Temporally ordered complete DAG



$$f(Y,A,L) = f(Y|L,A)\, f(A|L)\, f(L)$$

$$f(Y) = \Sigma_{a,l}\, f(Y|l,a)\, f(a|l)\, f(l)$$

Intervention DAG: Cut arrows into A and set A=a



$$f_a(Y,L) = f(Y|L,a)\, f(L)$$

$$f_a(Y) = \Sigma_l\, f(Y|l,a)\, f(l)$$

Put Y=1. We get sum over l E(Y|L=l,A=a) f(l)

G-formula and DAGS

Theorem: If $Y_a \amalg A \mid L$ *then* $f_a(y)=f(Y_a=y)$

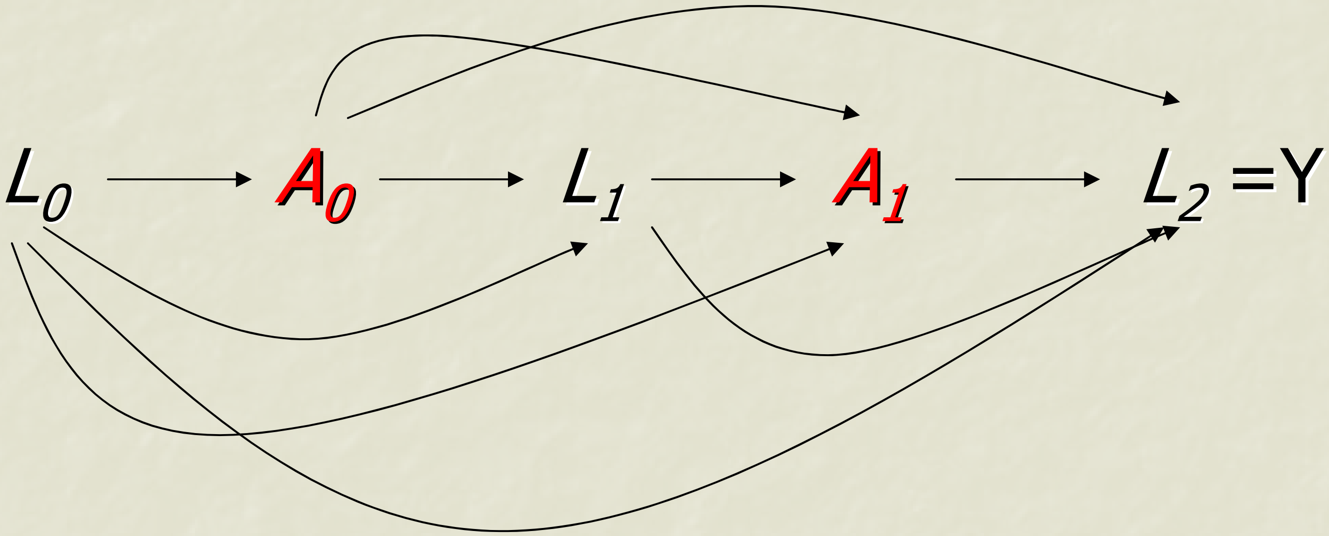
Proof:

$$\begin{aligned} f_a(y) &= \sum_l f(Y=y \mid l,a)\, f(l) \\ &= \sum_l f(Y_a=y \mid l,a)\, f(l) \\ &= \sum_l f(Y_a=y \mid l)\, f(l) \\ &= f(Y_a=y) \end{aligned}$$

old result if we put in y=1: G-formula gives mean of the counterfactual Ya. This generalizes to the probability of any level of Y

Time dependent treatments

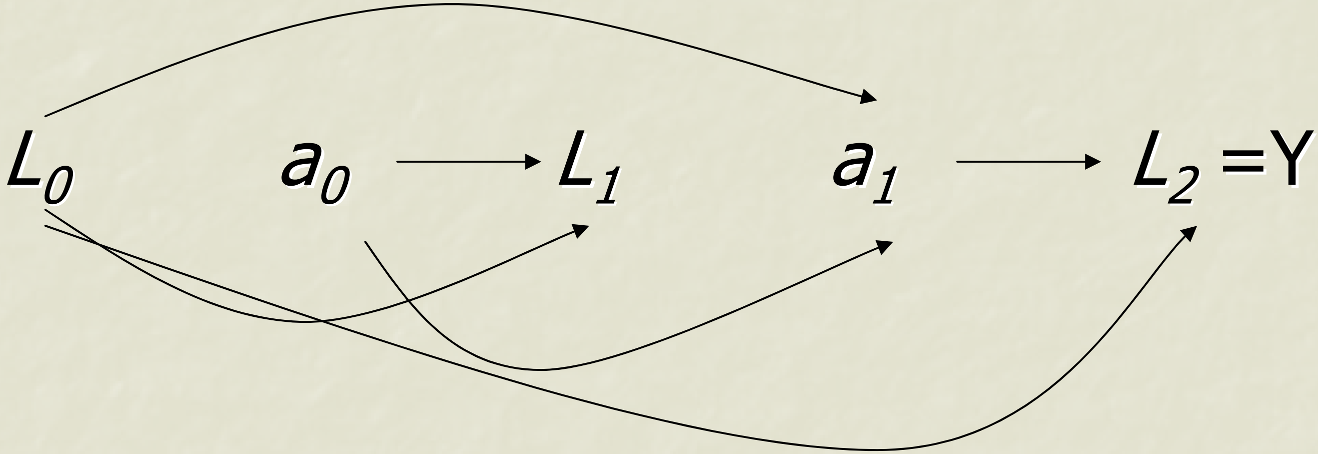
Temporally ordered complete DAG



$$f(L_2, A_1, L_1, A_0, L_0) = f(L_2 \mid A_1, L_1, A_0, L_0)\, f(A_1 \mid L_1, A_0, L_0)$$
$$f(L_1 \mid A_0, L_0)\, f(A_0 \mid L_0)\, f(L_0)$$

Time dependent treatments

Intervention DAG: cut arrows into A_0 and A_1 and set $A_0 = a_0$, $A_1 = a_1$



L_0 a_0 L_1 a_1 L_2 =Y

$$f_{a_0,a_1}(L_2,L_1,L_0) = f(L_2|a_1,L_1,a_0,L_0)f(L_1|a_0,L_0)f(L_0)$$

$$f_{a_0,a_1}(L_2) = \sum_{l_0,l_1} f(L_2|a_1,l_1,a_0,l_0)f(l_1|a_0,l_0)f(l_0)$$

G-formula and DAGS

Theorem: If there are no unmeasured confounders (sequential randomization holds) for the effect of Rx A_0 and A_1 on *Y, that is,*

$$Y_{a_0,a_1} \amalg A_0 | L_0 \;, \qquad Y_{a_0,a_1} \amalg A_1 | L_0, L_1, A_0$$

then

$$f_{a_0,a_1}(y) = f(Y_{a_0,a_1} = y)$$

in particular, if the outcome is binary,

$$f_{a_0,a_1}(1) = E(Y_{a_0,a_1})$$

Causal DAG

- Definition: A DAG is a causal DAG if for any ordered subset $(A_0, A_1, A_2, \ldots, A_K)$ of variables in the DAG

1. The causal effect for the remaining variables $(L_0, L_1, L_2, \ldots, L_{K+1})$ in the DAG is given by the G-formula

$$f_{a_0,a_1,a_2,\ldots,a_K}(l_0, l_1, \ldots, l_{K+1})$$

Or equivalently

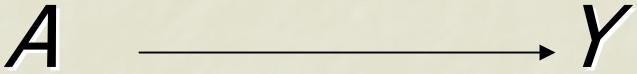
2. $\overline{L}_{a_0,a_1,\ldots,a_K} \amalg A_j | \overline{L}_j, \overline{A}_{j-1}$ for all j and all $a_0, a_1, \ldots, a_K$

(Notational convention)

$$\overline{U} = (U_0, \ldots, U_{K+1}) \qquad \overline{U}_j = (U_0, \ldots, U_j)$$

Example

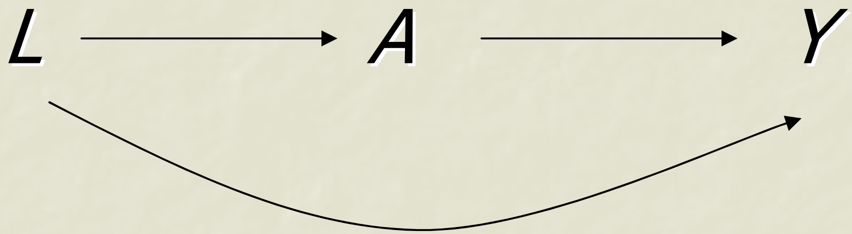
- If Y_a is NOT independent of A then



Is NOT a causal graph

- **Theorem:** *Given any two variables in a causal graph, all their common causes are in the graph (even if unmeasured)*

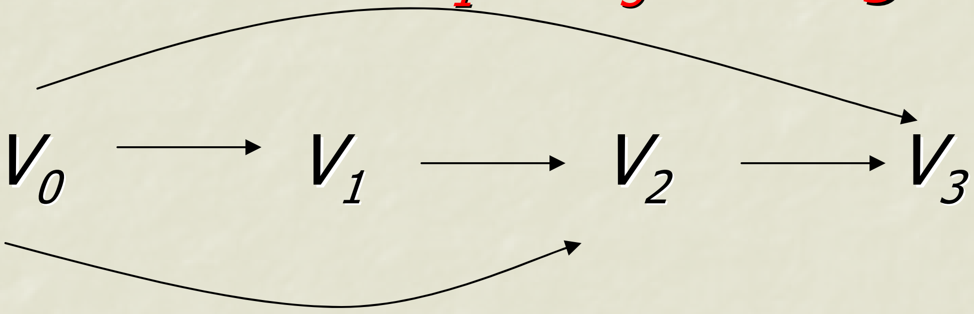
Example:



$$\{Y_a : a \in \mathcal{A}\} \amalg A | L$$

- *Causal DAGs need not be complete: Missing arrows means no direct effect*

Example: arrow from V_1 to V_3 missing



$$V_3 \amalg V_1 | V_0, V_2$$

If DAG is causal then V_1 does not have a direct effect on V_3 when we set V_0 *and* V_2, that is,

$f(V_{3,(v_0,v_1,v_2)} = v_3)$ does not depend on v_1

Proof: $f_{v_0,v_1,v_2}(v_3) = f(V_3 = v_3 | V_2 = v_2, V_1 = v_1, V_0 = v_0)$

$$= f(V_3 = v_3 | V_2 = v_2, V_0 = v_0)$$