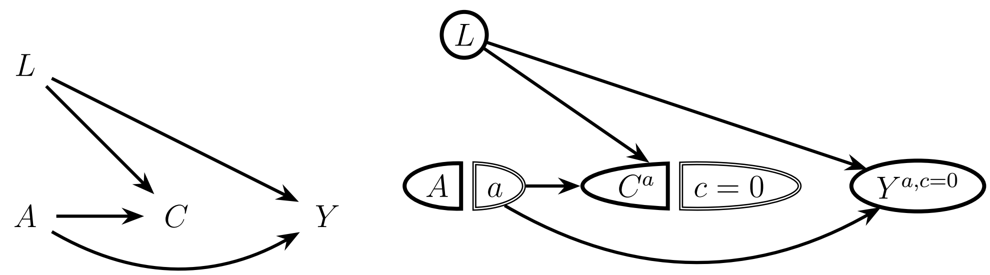
EXERCISES FOR RANDOMIZATION AND CAUSATION (MATH-336)

Exercise Sheet 6

Exercise 1 (Censoring)**.** Consider the DAG and SWIG below, reproduced from the lectures. Let $A, C, Y \in \{0,1\}$ be indicators of treatment, loss to follow-up and outcome respectively.



(a) Give one English sentence that explains the interpretation of $E[Y^{a,c=0}]$. Can we identify $E[Y^{a,c=1}]$ from the observed data distributions?
(b) Write down the positivity and conditional exchangeability assumption required to identify $E[Y^{a,c=0}]$.
(c) Find an identification formula for $E[Y^{a,c=0}]$.

Solution:

(a) $E[Y^{a,c=0}]$ is the expected (counterfactual) outcome under an intervention which sets treatment A to a and eliminates loss to follow-up (i.e. sets C to 0). In general, we cannot say anything about $E[Y^{a,c=1}]$, because we do not observe individuals after they are lost to follow-up.

Parts (b) and (c) are special cases of Exercise 2(e) of Exercise Sheet 5 with $L_0 = \emptyset$, $A_0 = A$ and $A_1 = C$.

(b) Conditional exchangeability for all $a \in \{0,1\}$:

$$Y^{a,c=0} \perp\!\!\!\perp A \; ,$$
$$Y^{a,c=0} \perp\!\!\!\perp C^a \mid A, L \; .$$

Positivity for all $a \in \{0,1\}$:

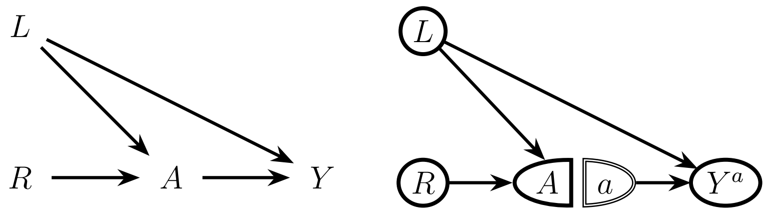
$$P(C=0 \mid L=l, A=a) > 0 \quad \text{whenever} \quad P(L=l, A=a) > 0 \; ,$$
$$P(A=a) > 0 \; .$$

(c) From Exercise 2(e) of Exercise Sheet 5, we know that

$$E[Y^{a,c=0}] = \sum_l E[Y \mid A=a, C=0, L=l] P(L=l) \; .$$

whenever consistency, conditional exchangeability and positivity hold.

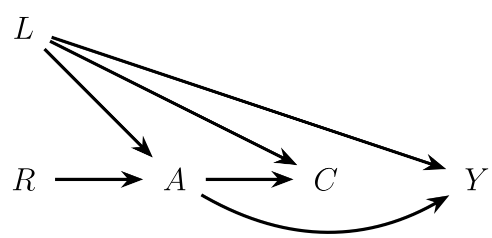
Exercise 2 (Imperfect adherence)**.** Consider a randomized trial where patients are assigned to one of two treatments $R \in \{0,1\}$ by randomization (flipping an unbiased coin) but do not necessarily adhere to their assigned treatment, such that their observed treatment level $A \in \{0,1\}$ may differ from R. Let L be a baseline covariate and let Y be the outcome. Suppose that all variables are binary, and assume that the causal model in the DAG and corresponding SWIG below are valid (The graphs are reproduced from lectures).



(a) (i) Write down an estimand for the per protocol effect (causal effect of A on Y). Write down the exchangeability conditions which allow us to identify the per protocol effect in a study with imperfect adherence.
(ii) Find an identification formula for this causal effect.

(b) (i) Write down an estimand for the intention-to-treat effect (the causal effect of R on Y). Write down the positivity and conditional exchangeability conditions which allow us to identify the intention-to-treat effect in a study with imperfect adherence (here, we assume no censoring)? Compare this to your answer in part (a)-(i)
(ii) Find an identification formula for this causal effect.

Next, we will consider a setting with imperfect adherence and losses to follow-up, depicted in the following DAG:



$\mathcal{G}$

(c) Write down the estimand for the causal effect of A on Y if we were to intervene to eliminate loss to follow-up, and draw the SWIG corresponding to this estimand. Write down the positivity and conditional exchangeability conditions which allow us to identify this estimand, and find an identification formula for this estimand.

Solution:

(a) (i) The estimand is $E[Y^a]$.
Conditional exchangeability and positivity for all $a \in \{0,1\}$:

$$Y^a \perp\!\!\!\perp A \mid L \ ,$$
$$P(A = a \mid L = l) > 0 \quad \text{whenever} \quad P(L > l) > 0 \ .$$

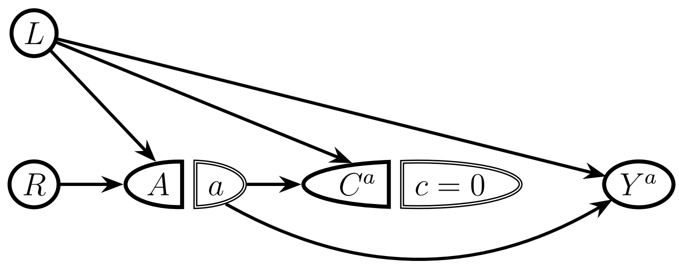
(ii) The identification formula is given by

$$E[Y^a] = \sum_y y \cdot P(Y^a = y) \overset{\text{LOTP}}{=} \sum_y \sum_l y \cdot P(Y = y \mid L = l) P(L = l)$$
$$\overset{Y^a \perp\!\!\!\perp A|L}{=} \sum_y \sum_l y \cdot P(Y = y \mid A = a, L = l) P(L = l)$$
$$\overset{\text{consistency}}{=} \sum_l E[Y \mid A = a, L = l] P(L = l)\ .$$

(b) (i) The estimand is given by $E[Y^r]$. We only require $Y^r \perp\!\!\!\perp R$ and $P(R = r) > 0$ for all $r \in \{0, 1\}$ to identify $E[Y^r]$. This is ensured to hold when R is randomized (assigned by flipping a coin), whereas $Y^a \perp\!\!\!\perp A \mid L$ is a strong assumption which requires detailed knowledge about the causes of failing to adhere to assigned treatment.

(ii) $E[Y^r] \overset{Y^r \perp\!\!\!\perp R}{=} E[Y^r \mid R = r] \overset{\text{consistency}}{=} E[Y \mid R = r]$.

(c) The estimand is given by $E[Y^{a,c=0}]$ and the desired SWIG $\mathcal{G}(a, c = 0)$ is given below. The identification conditions are :



$\mathcal{G}(a, c = 0)$

Conditional exchangeability for all $a \in \{0, 1\}$:

$$Y^{a,c=0} \perp\!\!\!\perp A \mid L,$$
$$Y^{a,c=0} \perp\!\!\!\perp C^a \mid A, L\ .$$

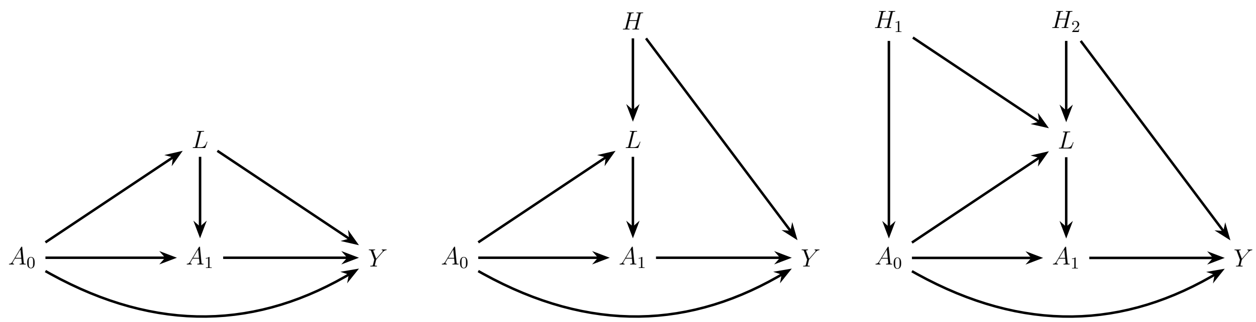
Positivity for all $a \in \{0, 1\}$:

$$P(C = 0 \mid L = l, A = a) > 0 \quad \text{whenever} \quad P(L = l, A = a) > 0\ ,$$
$$P(A = a|L = 0) > 0 \quad \text{whenever} \quad P(L = l) > 0.$$

The identification formula are identical to the one in Exercise 1(c).

Exercise 3 (Identification with hidden variables)**.** Consider another example of a sequentially randomized experiments, where the following measured variables are temporally (and topologically) ordered from left to right $\langle A_0, L, A_1, Y \rangle$, and any variable can depend on any other variable measured in its past.[1]

[1]The motivation is to show that we can identify causal effects in the presence of unmeasured variables. This follows straightforwardly from the identification theorem, which says that causal estimands are equal to the g-formula under certain conditions (positivity, conditional exchangeability and consistency). The idea is to check this manually in two special cases without using the identification theorem. As you can see, even these simple cases require some uses of algebra and independencies.



$\mathcal{G}_c$ $\mathcal{G}_H$ $\mathcal{G}_{H_1,H_2}$

(a) Draw the SWIG $\mathcal{G}_c(a_0, a_1)$. By assessing the conditional exchangeability assumptions for every path between treatments $A_0, A_1^{a_0}$ and outcome Y^{a_0,a_1}, convince yourself that $\mathcal{G}_c(a_0, a_1)$ satisfies the conditional exchangeability conditions

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

Use these conditions to show that

(1) $$P(Y^{a_0,a_1} = y) = \sum_l P(y \mid a_0, a_1, l)p(l \mid a_0) \ .$$

(b) Suppose next that there is a common cause H of L and Y. Draw the SWIG $\mathcal{G}_H(a_0, a_1)$. Convince yourself that it satisfies the conditional exchangeability conditions

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 \ ,$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, L^{a_0}, H \ .$$

Using these conditions, show that

$$P(Y^{a_0,a_1} = y) = \sum_l \sum_h P(y \mid a_1, a_0, l, h)p(l \mid a_0, h)p(h) \ .$$

(c) Draw the SWIG $\mathcal{G}_{H_1,H_2}(a_0, a_1)$ and convince yourself that it satisfies the conditional exchangeability conditions

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 \mid H_1 \ ,$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, L^{a_0}, H_1, H_2 \ .$$

Using these conditions, show that

$$P(Y^{a_0,a_1} = y) = \sum_l \sum_{h_1} \sum_{h_2} p(y \mid a_1, a_0, l, h_1, h_2)p(l \mid h_1, h_2, a_0)p(h_1)p(h_2) \ .$$

(d) By manipulating the conditional probabilities on the RHS of parts (b) and (c), show that both the right hand sides are equal to Eq. 1. Deduce that it is not necessary to measure H (or conversely H_1 and H_2) in order to identify $E[Y^{a_0,a_1}]$.[2]

[2] More broadly, it is not necessary to measure all causes of all variables in a causal model in order to identify causal effect, if positivity, conditional exchangeability and consistency hold. This is an important result, which tells us that we can study isolated parts of complex systems without knowing the full causal structure.

Hint: Use the laws of probability and independencies in the graphs $\mathcal{G}_H$ and $\mathcal{G}_{H_1,H_2}$ in order to express RHS on the form

$$\text{RHS}_{(\text{b})} = \sum_l \sum_h P(y, h \mid a_0, a_1, l) p(l \mid a_0)$$

and

$$\text{RHS}_{(\text{c})} = \sum_l \sum_{h_1} \sum_{h_2} P(y, h_1, h_2 \mid a_0, a_1, l) p(l \mid a_0)$$

in order to marginalize out the hidden variables by summing over them.

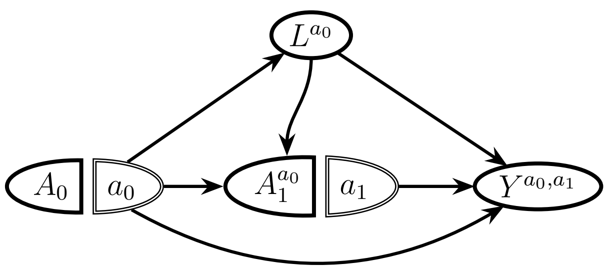
(e) Do the graphs in (b) and (c) satisfy the exchangeability conditions in part (a) (reproduced below)?

$$\begin{aligned} &Y^{a_0,a_1} \perp\!\!\!\perp A_0 \ , \\ &Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, L^{a_0} \ . \end{aligned}$$

Solution:

In what follows, we will use $(\cdot \perp\!\!\!\perp \cdot \mid \cdot)_{\mathcal{G}}$ to denote a conditional independence evaluated in a DAG, in order to distinguish it from a conditional independence evaluated in the corresponding SWIG.

(a) The SWIG $\mathcal{G}_c(a_0, a_1)$ is shown below. The exchangeability conditions are satisfied be-



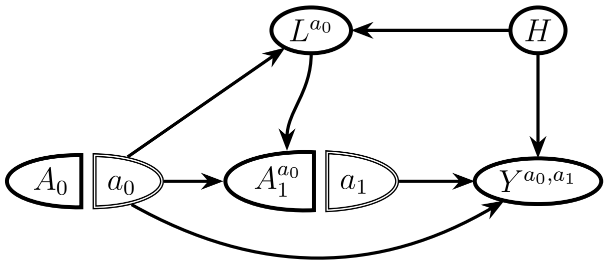
$\mathcal{G}_c(a_0, a_1)$

cause there are no opens paths between Y^{a_0,a_1} and A_0, or between Y^{a_0,a_1} and $A_1^{a_0}$ conditional on L^{a_0}, A_0. We thus have that

$$\begin{aligned} P(Y^{a_0,a_1} = y) &\overset{Y^{a_0,a_1} \perp\!\!\!\perp A_0}{=} P(Y^{a_0,a_1} = y \mid A_0 = 0) \\ &\overset{\text{LOTP}}{=} \sum_l P(Y^{a_0,a_1} = y \mid A_0 = a_0, L^{a_0} = l) P(L^{a_0} = l \mid A_0 = a_0) \\ &\overset{Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid L^{a_0}, A_0}{=} \sum_l P(Y^{a_0,a_1} = y \mid A_1^{a_0} = a_1, A_0 = a_0, L^{a_0} = l) P(L^{a_0} = l \mid A_0 = a_0) \\ &\overset{\text{consistency}}{=} \sum_l p(y \mid a_1, a_0, l) p(l \mid a_0) \ . \end{aligned}$$

(b) The SWIG $\mathcal{G}_H(a_0, a_1)$ is shown below. The exchangeability conditions are satisfied because there are no opens paths between Y^{a_0,a_1} and A_0, or between Y^{a_0,a_1} and $A_1^{a_0}$ conditional on L^{a_0}, A_0, H. We thus have that

$$P(Y^{a_0,a_1} = y) \overset{Y^{a_0,a_1} \perp\!\!\!\perp A_0}{=} P(Y^{a_0,a_1} = y \mid A_0 = 0)$$

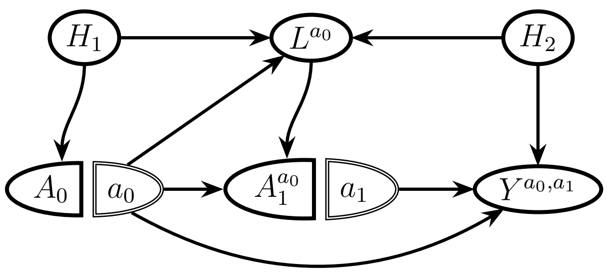


$\mathcal{G}_H(a_0, a_1)$

$$\overset{\text{LOTP}}{=} \sum_h P(Y^{a_0,a_1} = y \mid A_0 = a_0, H = h) \underbrace{P(H = h \mid A_0 = a_0)}_{\overset{(H \perp\!\!\!\perp A_0)_{\mathcal{G}}}{=} P(H=h)}$$

$$\overset{\text{LOTP}}{=} \sum_l \sum_h P(Y^{a_0,a_1} = y \mid L = l, A_0 = a_0, H = h)$$
$$\times P(L = l \mid A_0 = a_0, H = h) P(H = h)$$
$$\overset{Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid A_0, H, L^{a_0}}{=} \sum_l \sum_h P(Y^{a_0,a_1} = y \mid A_1^{a_0} = a_1, L = l, A_0 = a_0, H = h)$$
$$\times P(L = l \mid A_0 = a_0, H = h) P(H = h)$$
$$\overset{\text{consistency}}{=} \sum_l \sum_h P(y \mid a_1, a_0, l, h) p(l \mid a_0, h) p(h) \ .$$

(c) The SWIG $\mathcal{G}_{H_1,H_2}(a_0, a_1)$ is shown below. The exchangeability conditions are satisfied



$\mathcal{G}_{H_1,H_2}(a_0, a_1)$

because there are no opens paths between Y^{a_0,a_1} and A_0 conditional on H_1, or between Y^{a_0,a_1} and $A_1^{a_0}$ conditional on L^{a_0}, A_0, H_1, H_2. We thus have that

$$P(Y^{a_0,a_1} = y) \overset{\text{LOTP}}{=} \sum_{h_1} P(Y^{a_0,a_1} = y \mid H_1 = h_1) P(H_1 = h_1)$$
$$\overset{Y^{a_0,a_1} \perp\!\!\!\perp A_0 \mid H_1}{=} \sum_{h_1} P(Y^{a_0,a_1} = y \mid A_0 = a_0, H_1 = h_1) P(H_1 = h_1)$$
$$\overset{\text{LOTP}}{=} \sum_{h_2} \sum_{h_1} P(Y^{a_0,a_1} = y \mid H_2 = h_2, A_0 = a_0, H_1 = h_1)$$

$$
\begin{aligned}
&\times P(H_2 = h_2 \mid A_0 = a_0, H_1 = h_1) P(H_1 = h_1) \\
\overset{\text{LOTP}}{=} &\sum_l \sum_{h_2} \sum_{h_1} P(Y^{a_0,a_1} = y \mid L^{a_0} = l, H_2 = h_2, A_0 = a_0, H_1 = h_1) \\
&P(L^{a_0} = l \mid H_2 = h_2, A_0 = a_0, H_1 = h_1) \\
&\times \underbrace{P(H_2 = h_2 \mid A_0 = a_0, H_1 = h_1)}_{\overset{(H_2 \perp\!\!\!\perp (H_1, A_0))_{\mathcal{G}}}{=} P(H_2 = h_2)} P(H_1 = h_1).
\end{aligned}
$$

Finally, using $Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid H_2, H_1, L^{a_0}, A_0$, this is equal to

$$
\begin{aligned}
&\sum_l \sum_{h_2} \sum_{h_1} P(Y^{a_0,a_1} = y \mid A_1^{a_0} = a_1, L^{a_0} = l, H_2 = h_2, A_0 = a_0, H_1 = h_1) \\
&\qquad P(L^{a_0} = l \mid H_2 = h_2, A_0 = a_0, H_1 = h_1) \\
&\qquad \times P(H_2 = h_2) P(H_1 = h_1) \\
&\overset{\text{consistency}}{=} \sum_l \sum_{h_1} \sum_{h_2} p(y \mid a_1, a_0, l, h_1, h_2) p(l \mid h_1, h_2, a_0) p(h_1) p(h_2) \ .
\end{aligned}
$$

(d) We start with $\mathcal{G}_H$:

$$
\begin{aligned}
p(l \mid h, a_0) p(h) &\overset{(H \perp\!\!\!\perp A_0)_{\mathcal{G}}}{=} p(l \mid h, a_0) p(h \mid a_0) \\
&= p(l, h \mid a_0) \\
&= p(h \mid a_0, l) p(l \mid a_0) \\
&\overset{(H \perp\!\!\!\perp A_1 \mid L, A_0)_{\mathcal{G}}}{=} p(h \mid a_1, a_0, l) p(l \mid a_0).
\end{aligned}
$$

Plugging this into the identification formula in part (a) gives

$$
\begin{aligned}
P(Y^{a_0,a_1} = y) &= \sum_l \sum_h p(y \mid a_1, a_0, l, h) p(h \mid a_1, a_0, l) p(l \mid a_0) \\
&= \sum_l \sum_h p(y, h \mid a_0, a_1, l) p(l \mid a_0) \\
&= \sum_l P(y \mid a_0, a_1, l) p(l \mid a_0) \ .
\end{aligned}
$$

For $\mathcal{G}_{H_1,H_2}$, we begin by using graph independencies to show that

$$
\begin{aligned}
P(Y^{a_0,a_1} = y) &= \sum_l \sum_{h_1} \sum_{h_2} p(y \mid a_1, a_0, l, h_1, h_2) p(l \mid h_1, h_2, a_0) p(h_1) p(h_2) \\
&\overset{(Y \perp\!\!\!\perp L \mid A_1, A_0, H_1, H_2)_{\mathcal{G}}}{=} \sum_l \sum_{h_1} \sum_{h_2} p(y \mid a_1, a_0, h_1, h_2) p(l \mid h_1, h_2, a_0) p(h_1) p(h_2) \\
&\overset{(Y \perp\!\!\!\perp H_1 \mid A_1, A_0, H_2)_{\mathcal{G}}}{=} \sum_l \sum_{h_1} \sum_{h_2} p(y \mid a_1, a_0, h_2) p(l \mid h_1, h_2, a_0) p(h_1) p(h_2) \ .
\end{aligned}
$$

Next, we marginalize over L, H_1 and subsequently re-introduce them with different conditioning sets:[3]

$$\begin{aligned}
P(Y^{a_0,a_1}=y) &= \sum_{h_2} p(y \mid a_1,a_0,h_2)p(h_2)\sum_{h_1} p(h_1)\underbrace{\sum_{l} p(l \mid h_1,h_2,a_0)}_{=1} \\
&= \sum_{h_2} p(y \mid a_1,a_0,h_2)p(h_2)\underbrace{\sum_{h_1} p(h_1)}_{=1} \\
&= \sum_{h_2} p(y \mid a_1,a_0,h_2)p(h_2) \\
&= \sum_{h_2} p(y \mid a_1,a_0,h_2)p(h_2)\underbrace{\sum_{h_1} p(h_1 \mid a_0,h_2)}_{=1} \\
&= \sum_{h_2} p(y \mid a_1,a_0,h_2)p(h_2)\sum_{h_1} p(h_1 \mid a_0,h_2)\underbrace{\sum_{l} p(l \mid h_1,h_2,a_0)}_{=1} \\
&= \sum_{l}\sum_{h_1}\sum_{h_2} p(y \mid a_1,a_0,h_2)p(h_2)p(h_1 \mid a_0,h_2)p(l \mid h_1,h_2,a_0) \\
&\overset{(H_2 \perp\!\!\!\perp A_0)_{\mathcal{G}}}{=} \sum_{l}\sum_{h_1}\sum_{h_2} p(y \mid a_1,a_0,h_2)p(l \mid h_1,h_2,a_0)\underbrace{p(h_1 \mid a_0,h_2)p(h_2 \mid a_0)}_{=p(h_1,h_2 \mid a_0)} \;.
\end{aligned}$$

Using $(Y \perp\!\!\!\perp L \mid A_1, A_0, H_1, H_2)_{\mathcal{G}}$ and $(Y \perp\!\!\!\perp H_1 \mid A_1, A_0, H_2)_{\mathcal{G}}$ as we did above, we can re-introduce L and H_1 into the conditioning set for Y:

$$\begin{aligned}
P(Y^{a_0,a_1}=y) &= \sum_{l}\sum_{h_1}\sum_{h_2} p(y \mid a_1,a_0,h_1,h_2,l)p(l \mid h_1,h_2,a_0)p(h_1,h_2 \mid a_0) \\
&= \sum_{l}\sum_{h_1}\sum_{h_2} p(y \mid a_1,a_0,h_1,h_2,l)p(l,h_1,h_2 \mid a_0) \\
&= \sum_{l}\sum_{h_1}\sum_{h_2} p(y \mid a_1,a_0,h_1,h_2,l)p(h_1,h_2 \mid a_0,l)p(l \mid a_0) \\
&\overset{((H_1,H_2) \perp\!\!\!\perp A_1 \mid A_0,L)_{\mathcal{G}}}{=} \sum_{l}\sum_{h_1}\sum_{h_2} p(y \mid a_1,a_0,h_1,h_2,l)p(h_1,h_2 \mid a_0,a_1,l)p(l \mid a_0) \\
&= \sum_{l}\sum_{h_1}\sum_{h_2} p(y,h_0,h_1 \mid a_1,a_0,l)p(l \mid a_0) \\
&= \sum_{l} p(y \mid a_1,a_0,l)p(l \mid a_0) \;.
\end{aligned}$$

[3]There is also a simpler solution: first marginalize over L and H_1, then re-introduce L using the conditional probability $\sum_l p(l \mid a_0, h_2)$ (which is equal to 1) and finally re-introduce L into the conditioning set for Y. The resulting identification formula is identical to $\text{RHS}_{(b)}$ with H replaced by H_2. From here, we can prove the desired equality in the same way as for $\mathcal{G}_H$, using the corresponding graphical independencies.

(e) Yes, the graphs in (b) and (c) satisfy the exchangeability conditions in part (a), and are thus identified by the g-formula in part (a).

References