MAS 420: Crash Course in Good and Bad Controls

Adam Rohde
Department of Statistics
University of California, Los Angeles

October 25, 2022

Graphical Models and Good and Bad Controls

We'll discuss common settings when analysts need to decide whether or not to adjust for covariates in their identification and estimation of causal effects.

Useful resources

1. ($*$) A Crash Course in Good and Bad Controls - Cinelli et al (2022) [link]
2. A Graphical Catalog of Threats to Validity - Matthay and Glymour (2020) [link]
3. Single World Intervention Graphs (SWIGs) - Richardson and Robins (2013) [link]
4. Adjustment Criterion - Shpitser, VanderWeele, Robins (2012) [link]
5. Causal Inference: What If - Hernán and Robins (2020) [link]

Set up

Suppose we are interested in estimating the ATE of $X \in \{0,1\}$ on Y using covariate adjustment. That is, we want to identify the ATE using $Y_x \perp\!\!\!\perp X|Z$:

$$\begin{aligned}
\mathbb{E}[Y_1 - Y_0] &= \sum_z \mathbb{E}[Y_1 - Y_0|Z = z]p(Z = z) \\
&= \sum_z \left(\mathbb{E}[Y_1|Z = z] - \mathbb{E}[Y_0|Z = z]\right) p(Z = z) \\
&= \sum_z \left(\mathbb{E}[Y_1|X = 1, Z = z] - \mathbb{E}[Y_0|X = 0, Z = z]\right) p(Z = z) \text{ by } Y_x \perp\!\!\!\perp X|Z \\
&= \sum_z \left(\mathbb{E}[Y|X = 1, Z = z] - \mathbb{E}[Y|X = 0, Z = z]\right) p(Z = z) \text{ by ??}
\end{aligned}$$

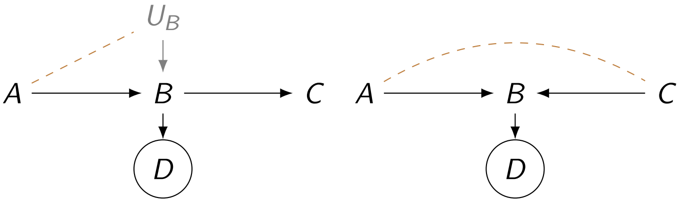
Say we want to use regression to estimate $\widehat{\mathbb{E}}[Y|X = 1, Z = z]$ and $\widehat{\mathbb{E}}[Y|X = 0, Z = z]$. So we want to estimate a regression function to predict the control potential outcomes and another to predict the treatment potential outcomes. But what covariates should we include as Z? How do we know whether some set of variables Z gives us $Y_x \perp\!\!\!\perp X|Z$?

Causal graphs

We will use graphs to do this. **Causal graphs capture our knowledge (and uncertainty) about how the variables we're studying relate causally. The most important assumptions are those that we do not draw!!** Basic building blocks:

- Chain: $A \rightarrow B \rightarrow C$. Relation between variables: $A \not\perp\!\!\!\perp C$ but $A \perp\!\!\!\perp C|B$
- Fork: $A \leftarrow B \rightarrow C$. Relation between variables: $A \not\perp\!\!\!\perp C$ but $A \perp\!\!\!\perp C|B$
- Collider: $A \rightarrow B \leftarrow C$. Relation between variables: $A \perp\!\!\!\perp C$ but $A \not\perp\!\!\!\perp C|B$

Also recall that conditioning on a descendant of a variable is the same as "partially" conditioning on it. In the first graph, $C \not\perp\!\!\!\perp A$ and $C \not\perp\!\!\!\perp A|D$ but the relationship changes conditional on D. In the second graph, $C \perp\!\!\!\perp A$ and $C \not\perp\!\!\!\perp A|D$.



Adjustment criterion

Definition (Adjustment criterion)

Z satisfies the adjustment criterion relative to X and Y in a causal graph G if

1. No element in Z lies on or is a descendant of a node on a causal path from X to Y.[a]
2. All non-causal paths from X to Y are blocked by Z.

[a] Elements of Z can be descendants of X if they are not on causal paths from X to Y.

Theorem

Assume the adjustment criterion holds for Z and (X, Y) in G. Then $Y_x \perp\!\!\!\perp X | Z$.[a]

[a] For every model inducing G.

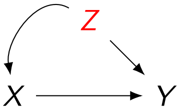
Single World Intervention Graphs (SWIGs)

SWIGs are a simple way to visualize potential outcomes on a graph.

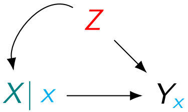
We turn a DAG into a SWIG by

1. Split the treatment (intervention) node X into two parts: an observed copy, X, and a copy that represents the value that we are intervening to set, x.
   - The observed copy, X, inherits all edges pointing **into** X in the DAG. Since this is the copy that represents the observed variable, it is still determined by all the parents it had in the DAG.
   - The interventional copy, x, inherits all edges pointing **out of** X in the DAG. Since this represents us intervening to set $X = x$, there are no parents for x in the SWIG.
2. Subscript all descendants of X in the DAG with a little x.

DAG (Model 1 from Crash Course)



SWIG for Y_x



Single World Intervention Graphs (SWIGs)

Since Y_x is on the SWIG, we can try to block the paths that connect X from Y_x in the graph as a way to get $Y_x \perp\!\!\!\perp X | Z$. We also check this against the conditions of the adjustment criterion. (Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG (Model 1 from Crash Course)

$$\begin{tikzcd} & Z \arrow[dl, bend right] \arrow[dr] & \\ X \arrow[rr] & & Y \end{tikzcd}$$

SWIG for Y_x

$$\begin{tikzcd} & Z \arrow[dl, bend right] \arrow[dr] & \\ X \mid x \arrow[rr] & & Y_x \end{tikzcd}$$

Z points to the observed copy of X, not the interventional copy.
The path $X \leftarrow Z \rightarrow Y_x$ needs to be blocked to get $Y_x \perp\!\!\!\perp X | Z$.
Here Z is a **good control**.

Practice

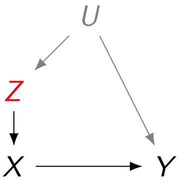
Our goals for working through the following slides are

- Practice applying the adjustment criterion
- See a variety of common types of good and bad "controls"
- Build some intuition for why the adjustment criterion says what it does by looking at SWIGs and seeing how potential outcomes relate to the treatment

Model 2 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

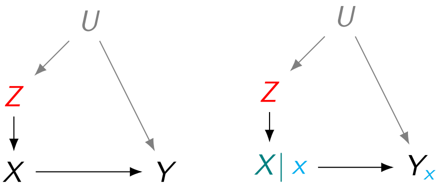
DAG



Model 2 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG SWIG for Y_x

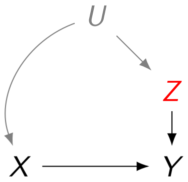


$X \leftarrow Z \leftarrow U \rightarrow Y_x$ makes $X \not\perp\!\!\!\perp Y_x$; conditioning on Z blocks this path.
Here Z is a **good control**.

Model 3 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X|Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

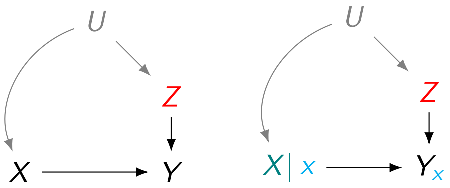
DAG



Model 3 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG SWIG for Y_x

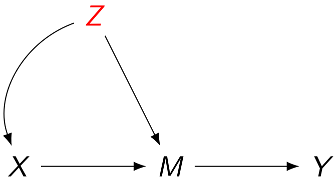


$X \leftarrow U \rightarrow Z \rightarrow Y_x$ makes $X \not\perp\!\!\!\perp Y_x$; conditioning on Z blocks this path.
Here Z is a **good control**.

Model 4 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

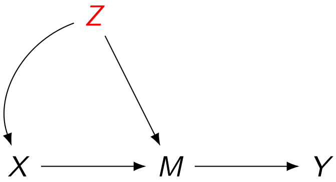
DAG



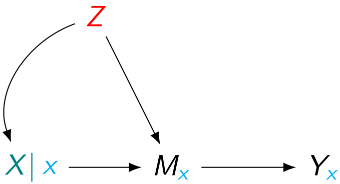
Model 4 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG



SWIG for Y_x

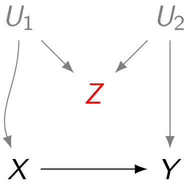


$X \leftarrow Z \rightarrow M_x \rightarrow Y_x$ makes $X \not\perp\!\!\!\perp Y_x$. But conditioning on Z blocks this path.
Here Z is a **good control**.

Model 7 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

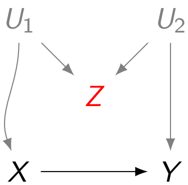
DAG



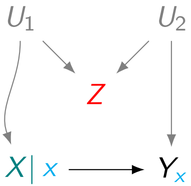
Model 7 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG



SWIG for Y_x

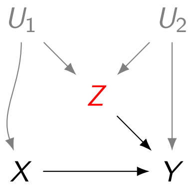


$X \leftarrow U_1 \rightarrow Z \leftarrow U_2 \rightarrow Y_x$ is blocked because Z is a collider. So $X \perp\!\!\!\perp Y_x$. Since Z is a collider, conditioning on it would open this path and so $X \not\perp\!\!\!\perp Y_x | Z$.
Here Z is a **bad control**.

Model 7 Variation from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

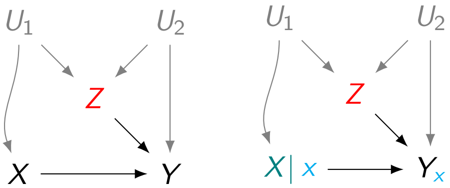
DAG



Model 7 Variation from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG SWIG for Y_x

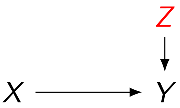


[1] $X \leftarrow U_1 \rightarrow Z \leftarrow U_2 \rightarrow Y_x$ is blocked because Z is a collider.
[2] $X \leftarrow U_1 \rightarrow Z \rightarrow Y_x$ is unblocked, however.
Conditioning on Z would block [2] but open [1].
So in this case both $X \not\perp\!\!\!\perp Y_x$ and $X \not\perp\!\!\!\perp Y_x | Z$.

Model 8 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG



Model 8 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG

$$\begin{tikzcd} & Z \arrow[d] \\ X \arrow[r] & Y \end{tikzcd}$$

SWIG for Y_x

$$\begin{tikzcd} & Z \arrow[d] \\ X \mid x \arrow[r] & Y_x \end{tikzcd}$$

There is no path from X to Y_x so $X \perp\!\!\!\perp Y_x$.
Since Z explains some of the residual variation in Y, conditioning on it could improve the precision of our estimates. Here Z is a **neutral control** but possibly good for precision.

Model 9 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

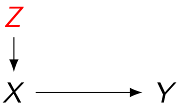
DAG

$$
\begin{tikzcd}
Z \arrow[d] & \\
X \arrow[r] & Y
\end{tikzcd}
$$

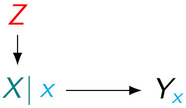
Model 9 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG



SWIG for Y_x

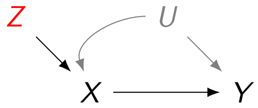


There is no path from X to Y_x so $X \perp\!\!\!\perp Y_x$.
Since Z explains some of the variation in X, conditioning on it could reduce the precision of our estimates. Here Z is a **neutral control** but possibly bad for precision.

Model 10 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X|Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

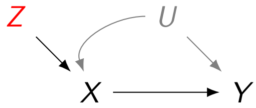
DAG



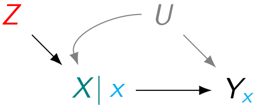
Model 10 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG



SWIG for Y_x



$X \leftarrow U \rightarrow Y_x$ means that $X \not\!\perp\!\!\!\perp Y_x$.
So an estimate of the effect of X on Y will be biased because of U.
It turns out that, *in linear models*, conditioning on Z can actually make the bias worse.
Here Z is a **bad control**. (See the appendix of the Crash Course for details.)

Model 11 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)? (Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG

$$X \longrightarrow Z \longrightarrow Y$$

Model 11 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)? (Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG

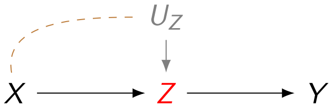
$$X \longrightarrow Z \longrightarrow Y$$

One way to think about this is that conditioning on Z blocks the path that we're trying to study and so we'd get biased estimates.
So Z is a **bad control**.

Model 11 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
Another view is that Z is a "virtual collider" between X and U_Z; conditioning on Z creates an association between X and U_Z, which creates a path between X and Y_x.

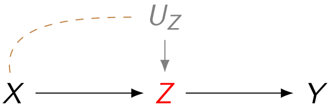
DAG



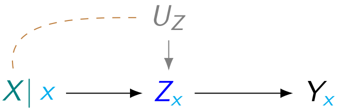
Model 11 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
Another view is that Z is a "virtual collider" between X and U_Z; conditioning on Z creates an association between X and U_Z, which creates a path between X and Y_x.

DAG



SWIG for Y_x



Z and Z_x are different variables. We only condition on Z not Z_x (a potential outcome).
$X - - U_Z \rightarrow Z_x \rightarrow Y_x$ means that $X \not\perp\!\!\!\perp Y_x | Z$.
Regardless of which perspective you prefer, Z is a **bad control**.

Model 12 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?

DAG

$$\begin{tikzcd} X \arrow[r] & M \arrow[r] \arrow[d] & Y \\ & Z & \end{tikzcd}$$

Model 12 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?

DAG

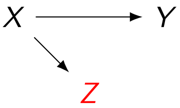
$$\begin{tikzcd} X \arrow[r] & M \arrow[r] \arrow[d] & Y \\ & \color{red}{Z} & \end{tikzcd}$$

One way to think about this is that conditioning on Z partially blocks the path that we're trying to study and so we'd get biased estimates. The more similar Z is to M, the closer we are to Model 11. So Z is a **bad control**.
We could also think about a "virtual collider" like in Model 11.

Model 14 from Crash Course

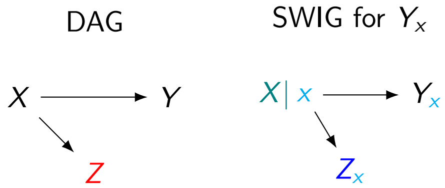
Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

DAG



Model 14 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

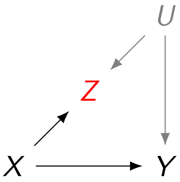


Z and Z_x are different variables. We only condition on Z not Z_x (a potential outcome).
There is no path from X to Y_x so $X \perp\!\!\!\perp Y_x$ and $X \perp\!\!\!\perp Y_x | Z$.
Z is a **neutral control** (possibly bad for precision).

Model 16 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

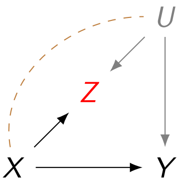
DAG



Model 16 from Crash Course

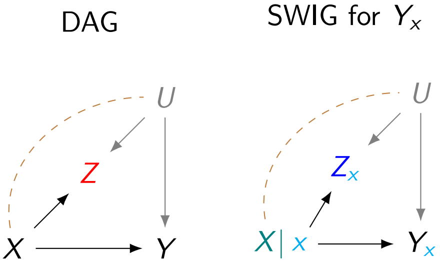
Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?

DAG



Model 16 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?



Since Z is a collider, conditioning on it opens $X - - U \to Y_x$ and so $X \not\perp\!\!\!\perp Y_x | Z$.
If we had not conditioned on Z then we wouldn't have this path and $Y_x \perp\!\!\!\perp X$.
Here Z is a **bad control**.

Model 18 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
(Are we conditioning on nodes on a causal path? Are we blocking all non-causal paths?)

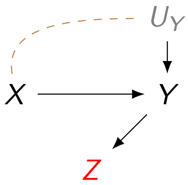
DAG

$$\begin{tikzcd}
X \arrow[rr] & & Y \arrow[dl] \\
& Z &
\end{tikzcd}$$

Model 18 from Crash Course

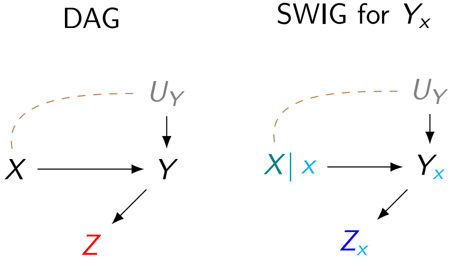
Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X | Z$)?
We again note that there is a virtual collider U_Y and conditioning on Z is conditioning on a descendant of that collider, which creates an association between X and U_Y.

DAG



Model 18 from Crash Course

Grey nodes are unobserved. Red nodes are the proposed conditioning variables.
Will conditioning on Z satisfy the adjustment criterion (and give $Y_x \perp\!\!\!\perp X|Z$)?
We again note that there is a virtual collider U_Y and conditioning on Z is conditioning on a descendant of that collider, which creates an association between X and U_Y.



$X - - U_Y \to Y_x$ means that $Y_x \not\perp\!\!\!\perp X$.
There is no observed variable that can block this path.
Z is a **bad control**.

Any remaining time

questions / break