Causal Inference
Discussion Assignment 2 (Study 2)

Alona Muzikansky, Jadey Wu, Alvaro J. Castro Rivadeneira, Jinghan Cui

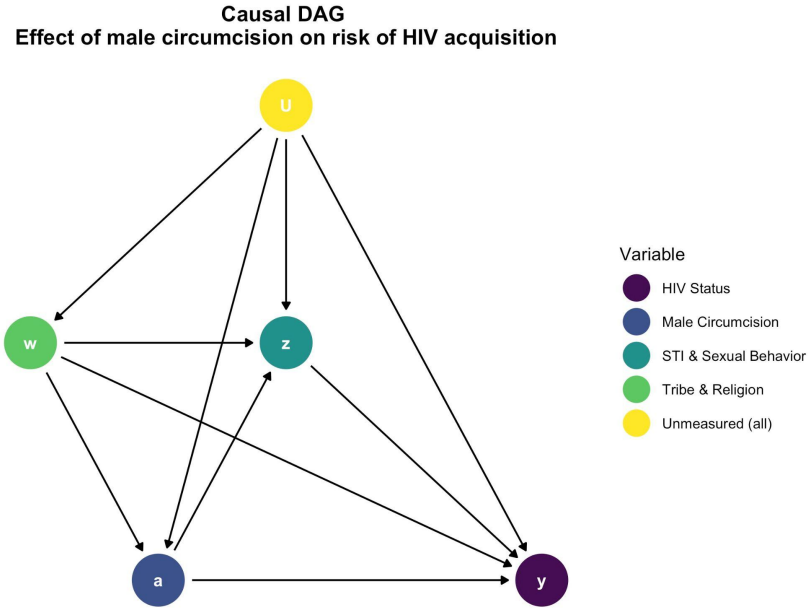
Step 3: Observed data and link to causal models

Observed data are $O = (W, A, Z, Y) \sim P_0$

W = {Tribe, Religion}
A = {Male circumcision}
Z = {Sexual behavior, STI}
Y = {HIV status}

The link between the SCM and the observed data is **n** identically distributed copies: $O_1, O_2, \ldots, O_n$ drawn from the probability distribution P_0 and assumed to be sampled from a system that is compatible with our SCM.

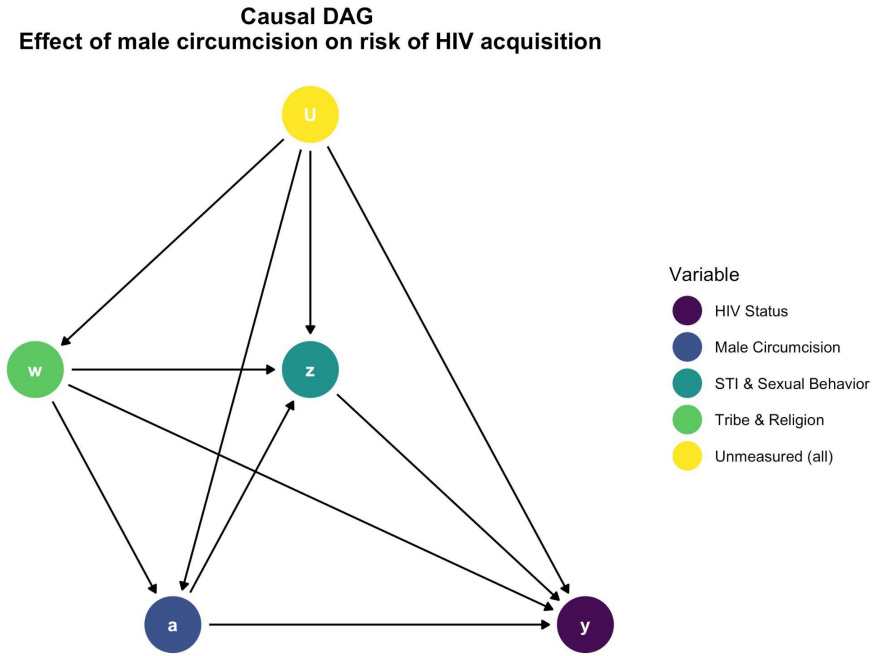
Our statistical model M is non-parametric and the SCM does not place any restrictions on M.



Step 4-5: Identifiability

Using the backdoor criteria, $\Psi^*(P^*)$ is not identifiable.

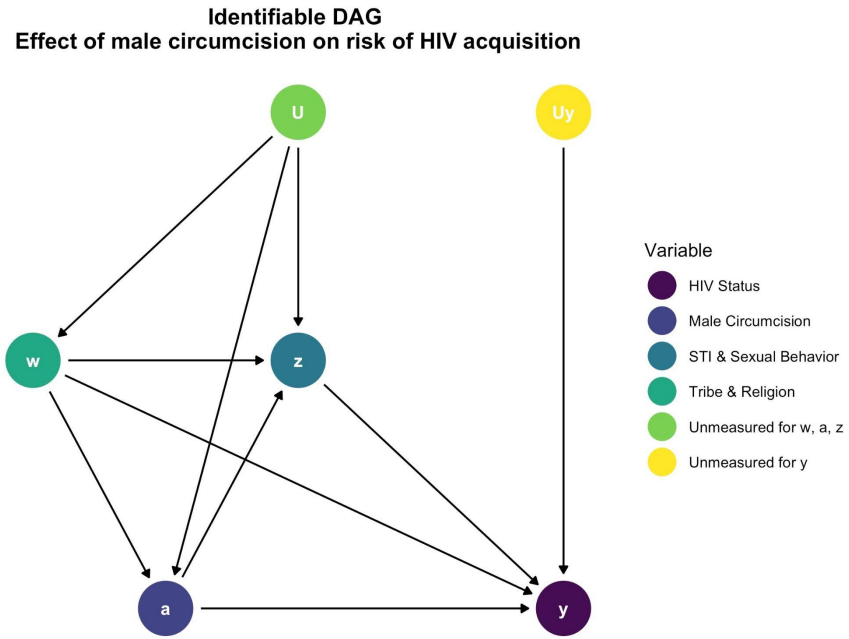
- {W,Z} can not block the path from U to A, so if we condition on {W,Z}, we still have a backdoor path from Y to U and then to A.



Step 4-5: Identifiability

To make $\Psi^*(P^*)$ identified, we can assume independence between U and U_Y or U and U_a.

- The independent assumption between U and U_Y is more plausible. Assumption of independent U_a is more possible under a randomized trial but this study is observational.
- And to assume independence between U_Y and U, we just need to assume any U_Y that is related to W and Z is measured and included in W and Z. Any left U_Y is independent of W and Z.



Step 4-5: Identifiability

- Additional measurements to make identifiability assumption more plausible:
  - Education
  - Socioeconomic status
  - Marital status
  - Sexual frequency
  - Condom usage
  - Sexual activity

Step 4-5: Identifiability

Specify the target parameter of the observed data distribution (ie, the statistical estimand)

$$\Psi(\mathbb{P}_0) = \sum_w [\mathbb{E}_0(Y|A=1, Z=z, W=w) - \mathbb{E}_0(Y|A=0, Z=z, W=w)]\mathbb{P}_0(Z=z|W=w))\mathbb{P}_0(W=w)$$

What is the relevant positivity assumption? Are you concerned about violations of the positivity assumption in your study?

The positivity assumption we have is:

$$min_{a \in \mathcal{A}} \mathbb{P}_0(A=a|Z=z, W=w) > 0$$

It means that for each group with certain sexual behavior within each tribe or religion, both circumcised and uncircumcised individuals must be observed. Violation of the positivity assumption is a concern with this study, as certain tribe/religion may have a preference toward circumcision. It may not be uncommon that all participants from the same tribe or religion are either circumcised or uncircumcised.

TABLE 1. Sociodemographic and Behavioral/HIV Risk Characteristics Among 1378 Men in the Kericho HIV Cohort Study*

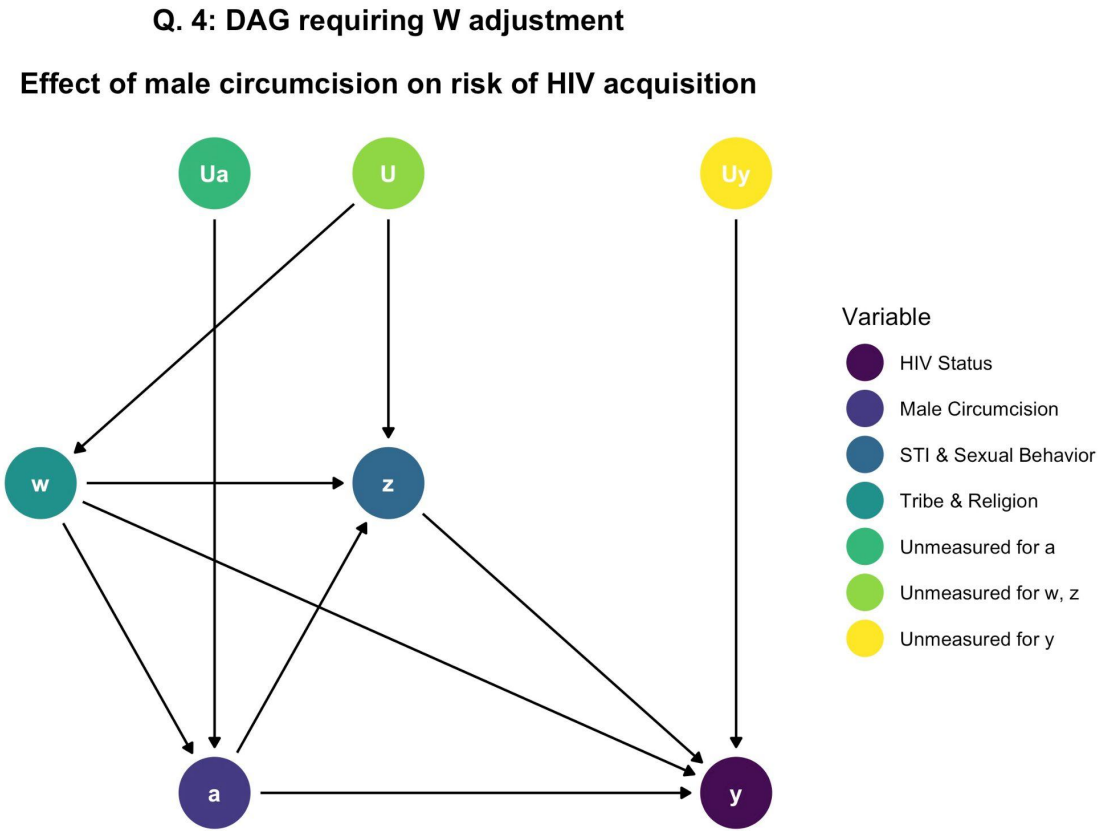
Feature	Total Subjects n	Total Subjects %	Uncircumcised (n = 270) n	Uncircumcised (n = 270) %	Circumcised (n = 1108) n	Circumcised (n = 1108) %	*P* Value
Age (y)							
18 to 24	343	24.9	70	25.9			
25 to 29	368	26.7	56	20.7			
30 to 34	243	17.6	40	14.8			
≥35	424	30.8	104	38.5			
Mean (standard deviation)	31.1 (8.8)		32.1 (9.6)				
Tribe							
Kalenjin	444	32.2	3	1.1	441	39.8	
Kisii	359	26.1	1	0.4	358	32.3	
Luhya	229	16.6	14	5.2	215	19.4	<0.001
Luo	231	16.8	200	74.1	31	2.8	
Others	115	8.3	52	19.3	63	5.7	
Religion							
Catholic	432	31.3	103	38.1	329	29.7	
Muslim	13	0.9	1	0.4	12	1.1	
Pentecostal	347	25.2	47	17.4	300	27.1	0.001
Protestant	383	27.8	79	29.3	304	27.4	
Traditional African	89	6.5	25	9.3	64	5.8	
Other	114	8.3	15	5.6	99	8.9	
Education							
High school or higher	663	48.8	106	40.3	557	50.8	
Primary	669	49.2	151	57.4	518	47.3	0.009
None	27	2.0	6	2.3	21	1.9	

These three groups are possible to violate the positivity assumption.

Question 4

Under what causal structure would *W* would satisfy the back door criterion, but (W,Z) would not?

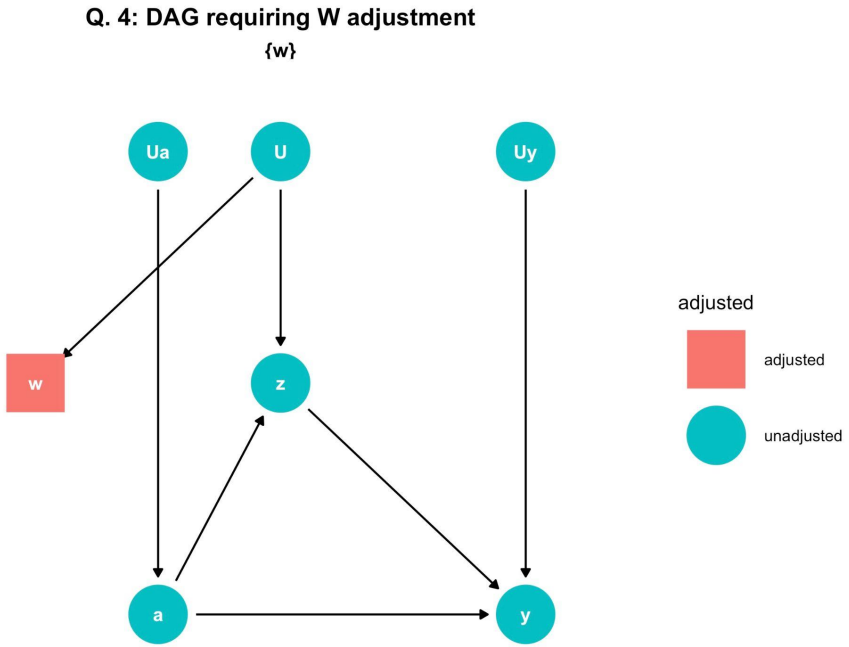
- When Ua and Uy are both independent with other U, conditioning on {W} alone satisfies the backdoor criterion, while conditioning on {W, Z} opens the path between Ua and U.



Question 4

- As shown in the graph, conditioning on W blocks all other path from A to Y. As Z is an instrumental variable, the point treatment effect can still be identified despite of the path from A to Z to Y.

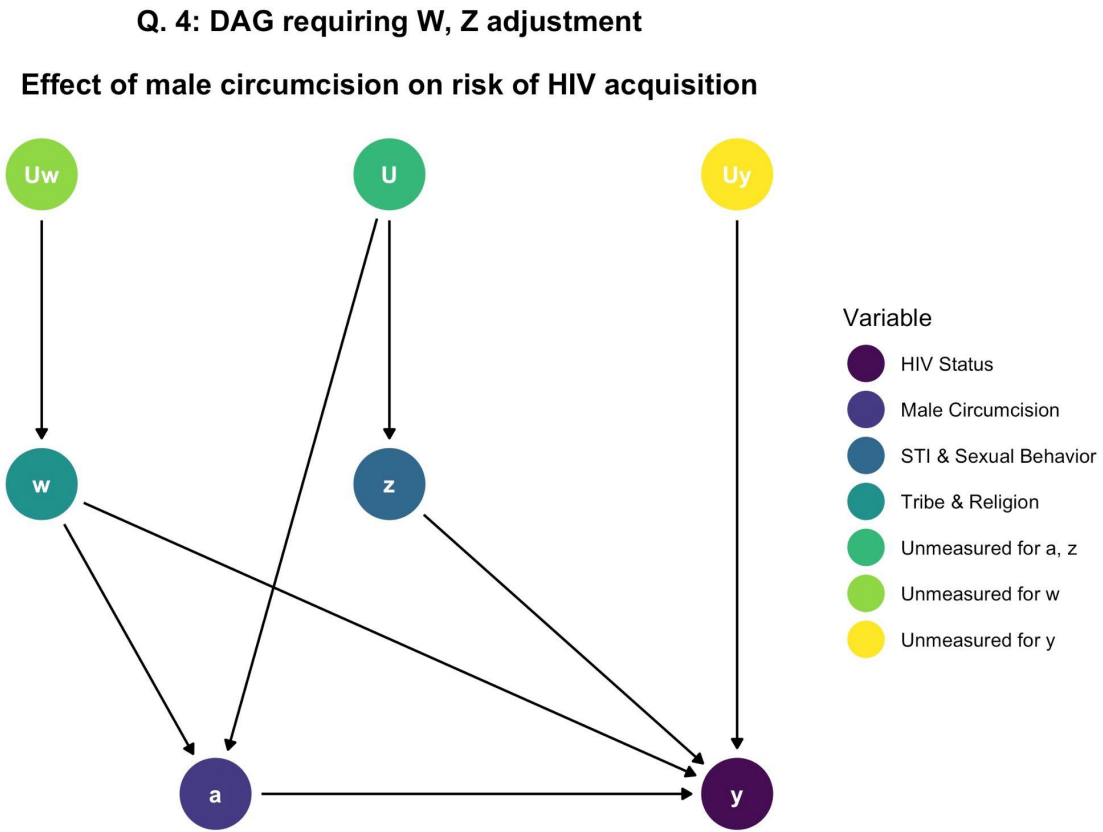
Under what causal structure would *W* would satisfy the back door criterion, but (W,Z) would not?



Question 4

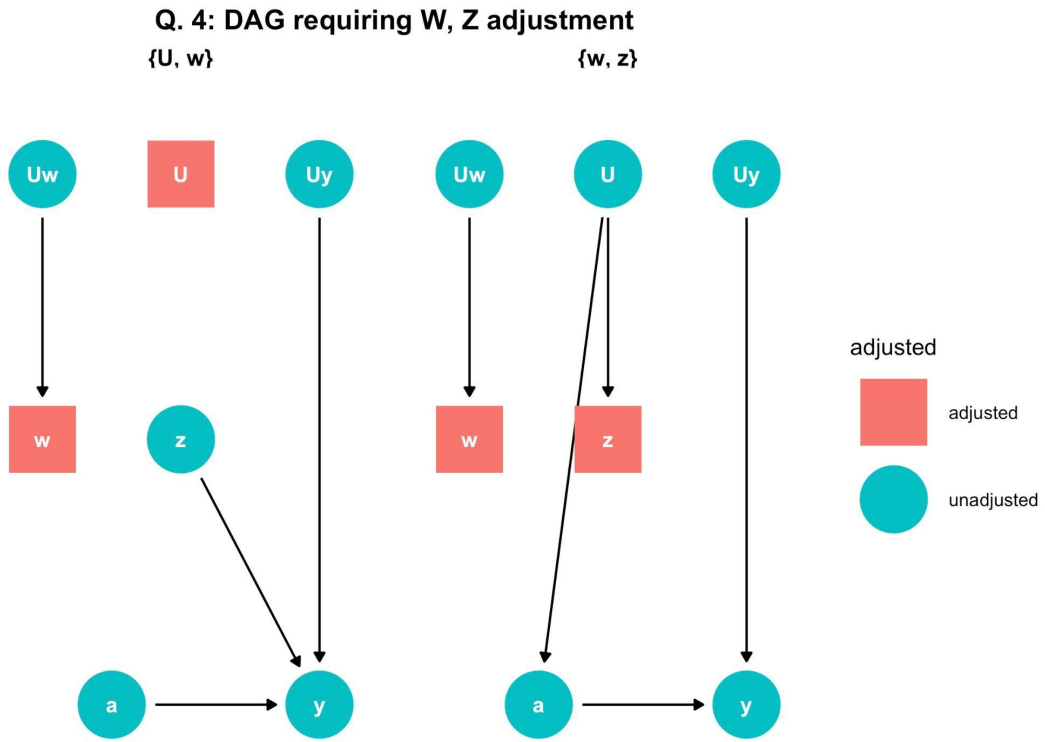
Under what causal structure would (W,Z) satisfy the back door criterion, but W alone would not?

- When U_W and U_Y are both independent with other U, conditioning on {W,Z} would satisfy the back door criterion but W alone would not.



Question 4

- To achieve identifiability in this causal structure, we need to either condition on {W,U} (DAG on the left) or {W,Z} (DAG on the right). Conditioning on W only will left a backdoor path Y-Z-U-a.
- However, U can not be conditioned, so only conditioning on {W,Z} can satisfy the backdoor criterion in this causal structure.



Thank you.

Questions?