Notes on Confounding

David Madigan
(based on Greenland & Morgenstern)

Confounding and Causality

- Confounding is a causal concept

	Population D		Population d	
Outcome	Drug (factual)	Not drug (counterfactual)	Drug (counterfactual)	Not drug (factual)
Y=1	30	20	30	10
Y=0	70	80	70	90
	***a*=0.3**	***b*=0.2**		***c*=0.1**

True causal effect = $a/b = 1.5$ or $a/(1-a) \div b/(1-b) = 1.71$
Estimated causal effect = $a/c = 3$ or $a/(1-a) \div c/(1-c) = 3.86$

- "The association in the combined D+d populations is confounded for the effect in population D"

Why does this happen?

- For confounding to occur there must be some characteristics/covariates/conditions that distinguish D from d.
- However, the existence of such factors does not in and of itself imply confounding.
- For example, D could be males and d females but it could still be the case that $b=c$.

Stratification can introduce confounding

	Population D		Population d	
Outcome	Drug (actual)	Not drug (counter)	Drug (counter)	Not drug (actual)
Y=1	30	20	30	20
Y=0	70	80	70	80
	a=0.3	**b=0.2**		**c=0.2**

True causal effect = a-b = 0.1
Estimated causal effect = a-c = 0.1
No confounding

Male

	Population D		Population d	
Outcome	Drug (actual)	Not drug (counter)	Drug (counter)	Not drug (actual)
Y=1	15	2	5	5
Y=0	35	8	65	15
	a=0.3	**b=0.2**		**c=0.25**

True = a-b = 0.1
Estimated = a-c = 0.05
Confounding

Female

	Population D		Population d	
Outcome	Drug (actual)	Not drug (counter)	Drug (counter)	Not drug (actual)
Y=1	15	18	25	15
Y=0	35	72	5	65
	a=0.3	**b=0.2**		**0.1875**

True = a-b =0.1
Estimated = a-c = 0.1125
Confounding

Non-Collapsibility without Confounding

Population D				
	Drug (factual)		Not drug (counterfactual)	
Covariate	Y=1	Y=0	Y=1	Y=0
Z=1	80	20	60	40
Z=0	40	60	20	80
Total	**120**	**80**	**80**	**120**

True causal effect | Z=1: 0.8 /0.2 ÷ 0.6/0.4 = 2.67
True causal effect | Z=0: 0.4 /0.6 ÷ 0.2/0.8 = 2.67
True causal effect ignoring Z: 0.6 /0.4 ÷ 0.4/0.6 = 2.25

Population d				
	Drug (counterfactual)		Not drug (factual)	
Covariate	Y=1	Y=0	Y=1	Y=0
Z=1			60	40
Z=0			20	80
Total			**80**	**120**

Estimated causal effect | Z=1: 0.8 /0.2 ÷ 0.6/0.4 = 2.67
Estimated causal effect | Z=0: 0.4 /0.6 ÷ 0.2/0.8 = 2.67
Estimated causal effect ignoring Z: 0.6 /0.4 ÷ 0.4/0.6 = 2.25

Collapsibility with Confounding

Population D				
	Drug (factual)		Not drug (counterfactual)	
Covariate	Y=1	Y=0	Y=1	Y=0
Z=1	80	20	60	40
Z=0	40	60	20	80
Total	**120**	**80**	**80**	**120**

True causal effect | Z=1: 0.8 /0.2 ÷ 0.6/0.4 = 2.67
True causal effect | Z=0: 0.4 /0.6 ÷ 0.2/0.8 = 2.67
True causal effect ignoring Z: 0.6 /0.4 ÷ 0.4/0.6 = 2.25

Population d				
	Drug (counterfactual)		Not drug (factual)	
Covariate	Y=1	Y=0	Y=1	Y=0
Z=1			60	40
Z=0			30	120
Total			**90**	**160**

Estimated causal effect | Z=1: 0.8 /0.2 ÷ 0.6/0.4 = 2.67
Estimated causal effect | Z=0: 0.4 /0.6 ÷ 0.2/0.8 = 2.67
Estimated causal effect ignoring Z: 0.6 /0.4 ÷ 0.36/0.64 = 2.67