

Mediating (post-treatment) variables may have a causal path to outcomes.

Child Care Example

Suppose a 3-year intervention with parents shows a 10 pt gain in kids IQs over control. Is it because of improved parenting?

Parenting Potential	Parenting Quality		Child's IQ		Proportion of sample
	after assigned to control	treated	after assigned to control	treated	
Poor either way	Poor	Poor	60	70	0.1
Good if treated	Poor	Good	65	80	0.7
Good either way	Good	Good	90	100	0.2

Note top and bottom lines show an improvement of 10. Middle line gets 15, so intervention is $5/15 = 1/3$ of the effect.

If we include both parenting and treatment in the model we get a negative treatment effect. Why? Treatment effect contrasts trt vs control when parenting is fixed.

Generate data by taking 20 kids (10 trt, 10 control, 2 line 1, 14 line 2, 4 line 3)

	poor	good	poor	good
control	64.38	90.00	8	2
treated	70.00	84.44	1	9

```
> iqs <- rep(c(60,65,90,70,80,100),c(1,7,2,1,7,2))
> trt <- factor(rep(0:1,each=10),labels=c("control","treated"))
> parenting <- factor(rep(c(0,0,1,0,1,1),c(1,7,2,1,7,2)),
  labels=c("poor","good"))
> coef(lm(iqs ~ trt + parenting))
```

(Intercept)	trttreated	parentinggood
65.18	-1.62	21.60

Treatment does NOT have a negative effect on IQ. Using the mediating variable "parenting" creates a problem. Randomization allows us to find causal effects of the treatment, not of mediators (unless you want to add a boat load of assumptions) Cannot answer: "What proportion of the treatment effect works through variable Z?"

Ideally we would divide the study into the 3 levels in the table above and study each in turn. (Line 2 is of the most interest and gives strongest results). BUT these strata are unknown.

Can we ever find substitutes – principal strata? Need assumptions akin to those for ignorability.

More in Chapter 10.

Still a problem, now doubles the ignorability problem.

Wells again

$\Pr(\text{switch}) =$

$\text{logit}^{-1}(-0.21 - .90 \cdot \text{dist100} + 0.47 \cdot \text{arsenic} + 0.17 \cdot \text{educ4})$

Can we interpret it causally for distance, arsenic level and education?

Only if we are changing one while holding the other 2 fixed. Can we do that?

Think about digging a new well at distance x from a house with an arsenic problem. What would make them switch to it? Can we model cost and benefit?