Effect measures

Stijn Vansteelandt

Ghent University, Belgium
London School of Hygiene and Tropical Medicine, U.K.

Background

- So far, we have been speaking rather loosely about causal effect and direct effect.
- Once we wish to report effect sizes, we need to be more precise.
- The use of counterfactuals forces one to be clear and precise.

Counterfactuals

- Suppose that we are interested in the causal effect of smoking A during pregnancy on stillbirth Y.
- Suppose for Emma, we could observe her baby's mortality status $Y(1)$ if she were to smoke during pregnancy.
- Suppose that we could also observe her baby's mortality status $Y(0)$ if she never smoked during pregnancy, all other things being the same.
- $Y(0)$ and $Y(1)$ are referred to as counterfactual or potential outcomes.

Causal effects

- If $Y(1) = 1$ and $Y(0) = 0$,
  then Emma's smoking would cause her baby to die.
- Her individual causal effect of smoking on her baby's mortality is

$$Y(1) - Y(0)$$

  It is unobservable.
- The population causal effect (Hernán, 2004) is

$$E\{Y(1) - Y(0)\}.$$

  It can be identified under certain assumptions (e.g. randomization).

Causal effect versus association

Mother	A	Y	$Y(0)$	$Y(1)$
Emma	1	1	1	1
Anna	1	1	1	1
Mary	1	0	0	0
Stephanie	0	0	0	0
Andrea	0	0	0	0
Kathy	0	1	1	1

$$E\{Y(1)\} - E\{Y(0)\} = E\{Y(1) - Y(0)\} = 0$$

versus

$$E(Y|A=1) - E(Y|A=0) = 1/3$$

Multiple versions of treatment

Nutrition Research Reviews (2009), **22**, 93–108
doi:10.1017/S0954422409990035

A review and meta-analysis of the effect of weight loss on all-cause mortality risk

Mary Harrington[1], Sigrid Gibson[2] and Richard C. Cottrell[3]*

[1]*The Sugar Bureau, London WC2B 5JJ, UK*
[2]*Sig-Nurture Ltd, Guildford, Surrey GU1 2TF, UK*
[3]*World Sugar Research Organisation, London SW1V 3LX, UK*

What is meant by the effect of weight loss on mortality?

Linear regression

- If L is sufficient to adjust for confounding and

$$E(Y|A, L) = \beta_0 + \beta_1 A + \beta_2 L$$

  then β_1 can be interpreted
  as a conditional causal effect or subgroup effect

$$E\{Y(1) - Y(0)|L\} = \beta_1$$

- Because all subgroups have the same effect,
  it can also be interpreted
  as a marginal causal effect or population-averaged effect

$$E\{Y(1) - Y(0)\}$$

Logistic regression

- If L is sufficient to adjust for confounding and

$$\mathrm{logit} E(Y|A, L) = \beta_0 + \beta_1 A + \beta_2 L$$

then $\exp(\beta_1)$ can be interpreted
as a conditional causal effect or subgroup effect

$$\frac{\text{odds}\{Y(1) = 1|L\}}{\text{odds}\{Y(0) = 1|L\}}$$

- Due to noncollapsibility of the odds ratio,
it generally differs from the marginal causal effect

$$\frac{\text{odds}\{Y(1) = 1\}}{\text{odds}\{Y(0) = 1\}} \neq \exp(\beta_1)!$$

Simulation experiment

```r
n <- 1e6
a <- rbinom(n,1,0.5)
l <- rnorm(n)
y <- rbinom(n, 1, expit(a + l))
```

Is there confounding of the effect of A on Y?
Draw a causal diagram corresponding to this data-generating mechanism.

Conditional causal effect

```
> mod1 <- glm(y ~ a + l, family=binomial)
> summary(mod1)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.001252   0.003109  -0.403    0.687
a            1.003000   0.004644 215.979   <2e-16 ***
l            1.000094   0.002715 368.306   <2e-16 ***
```

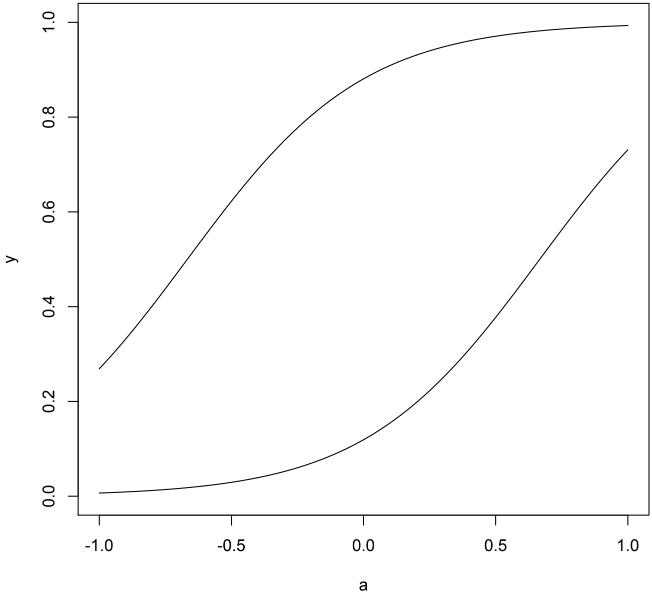
$$\frac{\text{odds}\left\{Y(1)=1|L\right\}}{\text{odds}\left\{Y(0)=1|L\right\}} = \frac{\text{odds}\left(Y=1|A=1,L\right)}{\text{odds}\left(Y=1|A=0,L\right)} = \exp(1) = 2.72$$

Marginal causal effect

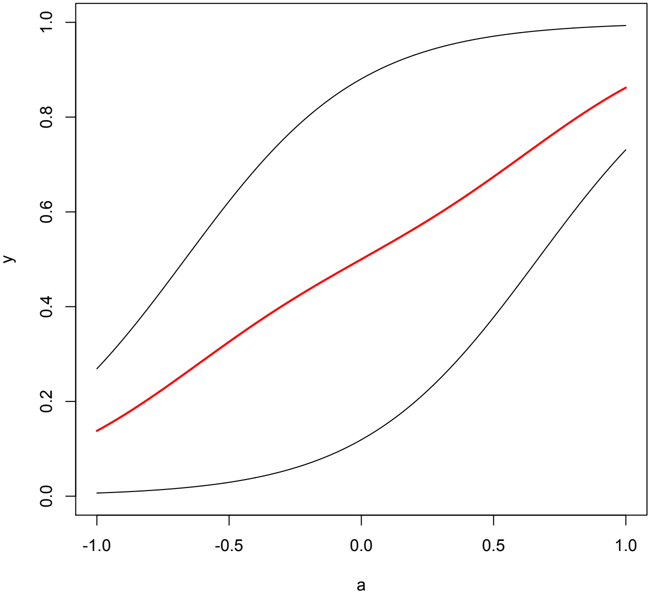
```
> mod2 <- glm(y ~ a, family=binomial)
> summary(mod2)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.001905   0.002826  -0.674      0.5
a            0.837964   0.004181 200.406   <2e-16 ***
```

$$\frac{\text{odds}\left\{Y(1)=1\right\}}{\text{odds}\left\{Y(0)=1\right\}} = \frac{\text{odds}\left(Y=1|A=1\right)}{\text{odds}\left(Y=1|A=0\right)} = \exp(0.84) = 2.31$$

Conditional or subgroup effects



Marginal or population-averaged effect is diluted



Case study: Westphalian Stroke Registry

- Includes all patients treated in Northwestern Germany for stroke symptoms, admitted to 42 participating hospitals.
  (Kurth et al., *AJE* 06)
- 8208 ischemic stroke patients, between 2000 and 2001.
- Goal: Effect of tissue plasminogen activator on death.

Case study: Westphalian Stroke Registry

Method	No.	OR	95% CI
No adjustment	6269	3.35	2.28 - 4.91
Ordinary regression	6269	1.93	1.22 - 3.06
Matching	406	1.17	0.68 - 2.00
IPTW	6269	10.77	2.47 - 47.04

What explains the differences?

Different methods infer different effect measures

- Ordinary regression infers

$$\frac{\text{odds}\{Y(1) = 1|L\}}{\text{odds}\{Y(0) = 1|L\}}$$

- Matching infers

$$\frac{\text{odds}\{Y(1) = 1|A = 1\}}{\text{odds}\{Y(0) = 1|A = 1\}}$$

- IPTW infers

$$\frac{\text{odds}\{Y(1) = 1\}}{\text{odds}\{Y(0) = 1\}}$$

These effects are of a different magnitude

	Treated ($n = 212$)			Not Treated ($n = 6057$)			
Percent.	PS	No.	%	PS	No.	%	OR
99-100	0.58	36	8.3	0.55	26	26.9	0.25
95-99	0.31	73	17.8	0.29	178	15.2	1.21
90-95	0.14	55	14.6	0.14	258	7.4	2.14
75-90	0.059	31	9.7	0.046	910	9.0	1.08
50-75	0.012	10	40	0.0084	1558	5.6	11.3
25-50	0.0017	5	40	0.0014	1561	3.5	18.6
10-25	0.0004	2	50	0.00027	940	3.8	25.1
5-10	0	0	0	$6.6\ 10^{-5}$	313	1.9	
1-5	0	0	0	$2.7\ 10^{-5}$	251	3.2	
0-1	0	0	0	$7\ 10^{-6}$	62	1.6	
Overall	0.25	212	16.0	0.0262	6057	5.4	3.35

Summary

Take home message: intervention effects can be defined in many ways

- Standard regression procedures infer conditional effects.
- But interest lies often in marginal effects,
  or effects in the (un)treated.
- Counterfactuals force us to be explicit
  about the meaning of an intervention's effect.

Summary

**Take home message: adding variables to a regression model can change the magnitude of the effects,
even when the exposure is randomly assigned**

$$\begin{tikzcd} & L \arrow[dr] & \\ A \arrow[rr] & & Y \end{tikzcd}$$

- e.g. due to effect modification or non-collapsibility.
- Standard model building procedures based on evaluating changes-in-coefficients are thus fallible.
- Effect sizes can be difficult to compare between studies!

  (watch out for meta-analyses)

References

Greenland S, Robins JM and Pearl J. Confounding and collapsibility in causal inference. Statistical Science 1999; **14**, 29-46.

Hernan MA. A definition of causal effect for epidemiological research. *Journal of Epidemiology and Community Health* 2004; **58**, 265-271.

Vansteelandt S and Keiding N. Invited Commentary: G-Computation - Lost in Translation? *Am J Epidem* 2011;**173**:739-742.

Methods in Research on Research

A presentation delivered at the

first MiRoR training event
October 19-21, 2016
Ghent, Belgium

This project has received funding from the EU Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Grant Agreement #676207

www.miror-ejd.eu

@MirorProject