

# Marginal odds ratios

What they are, how to compute them,  
and why applied researchers might want to use them

Ben Jann and Kristian Bernt Karlson

2022 Swiss Stata Conference  
University of Bern, November 18, 2022

# Outline

- 1 Background
- 2 Marginal (log) odds ratios
- 3 Estimation
- 4 Example
- 5 Discussion

Acknowledgements: This presentation is based on work which is part of a project by Kristian Bernt Karlson that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 851293).

# Background

- Odds ratios form the backbone of much quantitative research in social sciences and epidemiology.
- Close to a hallmark of these disciplines!
- But: Falling out of favor!
  - ▶ Magnitude of odds ratios depends on unmeasured covariates orthogonal to the predictor of interest.
  - ▶ Noncollapsibility (rescaling bias).
  - ▶ Invalidates cross-model and subgroup coefficient comparisons.

# Background

- Solutions?
- KHB for cross-model comparisons (Karlson et al. 2012)
- Compare sign not magnitude
- Average marginal effects (AME)
- Linear probability models (LPM)

# Background

- At least in social sciences, AMEs are now the standard.
- But this might be throwing out the baby with the bathwater, because ...
  - ... magnitudes depend on the margin
  - ... AMEs focus on absolute probability differences, not relative differences, which is key to much theory and research.

# Background

- What we suggest:

Use **marginal (log) odds ratios**, which . . .

. . . behave like AME but retain the (relative) odds ratio interpretation!

- ✓ unaffected by noncollapsibility
- ✓ an average effect (population-averaged)
- ✓ comparable across populations/studies

- 1 Background
- 2 Marginal (log) odds ratios
- 3 Estimation
- 4 Example
- 5 Discussion

# Marginal odds ratio

- Following Zhang (2008) and Daniel et al. (2021) we use potential outcomes notation to define the marginal odds ratio.
- $Y_t$ : Potential outcome that would realize if treatment  $T$  was set to level  $t$  by manipulation (i.e., without changing anything else).
- We focus on *binary* outcomes only, that is,  $Y_t \in \{0, 1\}$  (failure or success).
- Thus:

$\Pr(Y_t = 1) = E[Y_t]$  is the (marginal) probability that  $Y_t$  will be equal to 1 (probability of success).



## Marginal odds ratio

- Consider a binary treatment  $T \in \{0, 1\}$ .
- The marginal odds ratio (MOR) of the alternative treatment ( $T = 1$ ) versus the standard treatment ( $T = 0$ ) is defined as

$$\text{MOR} = \frac{v[\Pr(Y_1 = 1)]}{v[\Pr(Y_0 = 1)]} = \exp\{\ln v[\Pr(Y_1 = 1)] - \ln v[\Pr(Y_0 = 1)]\}$$

where  $v(p) = p/(1 - p)$  (odds) and  $\ln v(p) = \ln(p/(1 - p))$  (log odds).

- Interpretation of MOR: The ratio of the odds of success if everyone would receive the alternative treatment versus the odds of success if everyone would receive the standard treatment (assuming that there are no general equilibrium effects, i.e., SUTVA holds).

“Marginal” refers to how a predictor affects the “marginal distribution” of an outcome (i.e., not to a marginal change in a predictor). “Unconditional” would be another term but we use “marginal” because the term is established in the literature (Stampf et al. 2010; Karlson, Popham, and Holm 2021).

## Adjusting for covariates

- The probability of success may not only depend on  $T$ , but also on other factors  $\mathbf{X}$ .
- Assume that  $\mathbf{X}$  has a specific distribution in the population and let  $\Pr(Y_t = 1|\mathbf{X} = \mathbf{x}) = E[Y_t|\mathbf{X} = \mathbf{x}]$  be the conditional success probability given  $\mathbf{X} = \mathbf{x}$ .
- By the law of iterated expectations,

$$\Pr(Y_t = 1) = E_{\mathbf{X}}[\Pr(Y_t = 1|\mathbf{X} = \mathbf{x})]$$

where  $E_{\mathbf{X}}$  is the expectation over the distribution of  $\mathbf{X}$ .

## Adjusting for covariates

- The marginal odds ratio, adjusting for  $\mathbf{X}$ , can then be written as

$$\begin{aligned}\text{MOR} &= \frac{\nu\{E_{\mathbf{X}}[\Pr(Y_1 = 1|\mathbf{X} = \mathbf{x})]\}}{\nu\{E_{\mathbf{X}}[\Pr(Y_0 = 1|\mathbf{X} = \mathbf{x})]\}} \\ &= \exp(\ln \nu\{E_{\mathbf{X}}[\Pr(Y_1 = 1|\mathbf{X} = \mathbf{x})]\} - \ln \nu\{E_{\mathbf{X}}[\Pr(Y_0 = 1|\mathbf{X} = \mathbf{x})]\})\end{aligned}$$

- We term this the *adjusted* MOR.
- Note:
  - ▶ The adjusted MOR is the same as the unadjusted MOR by definition (i.e., same estimand)!
  - ▶ However, estimation based on the adjusted MOR formulation can be used to address confounding bias in observational data. It can also be used to increase efficiency in analysis of randomized experiments.

## Continuous treatment

- In case of a continuous treatment, the MOR may depend on the level of the treatment (i.e., MOR may not be constant).
- We define the level-specific marginal log odds ratio as the derivative of the marginal log odds by the treatment:

$$\begin{aligned}\ln \text{MOR}(t) &= \lim_{\epsilon \rightarrow 0} \frac{\ln v[\Pr(Y_{t+\epsilon} = 1)] - \ln v[\Pr(Y_t = 1)]}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\ln v\{E_{\mathbf{X}}[\Pr(Y_{t+\epsilon} = 1 | \mathbf{X} = \mathbf{x})]\} - \ln v\{E_{\mathbf{X}}[\Pr(Y_t = 1 | \mathbf{X} = \mathbf{x})]\}}{\epsilon}\end{aligned}$$

- We can then obtain the average MOR by integrating over the distribution of  $T$ :

$$\overline{\text{MOR}} = \exp\{E_T[\ln \text{MOR}(t)]\}$$

## Continuous treatment

- Another possibility is to integrate over  $T$  when obtaining the population-averaged probabilities, that is,

$$\begin{aligned}\ln \text{MOR}' &= \lim_{\epsilon \rightarrow 0} \frac{\ln v\{E_T[\Pr(Y_{t+\epsilon} = 1)]\} - \ln v\{E_T[\Pr(Y_t = 1)]\}}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\ln v\{E_{T,\mathbf{X}}[\Pr(Y_{t+\epsilon} | \mathbf{X} = \mathbf{x})]\} - \ln v\{E_{T,\mathbf{X}}[\Pr(Y_t | \mathbf{X} = \mathbf{x})]\}}{\epsilon}\end{aligned}$$

- This corresponds to the marginal odds ratio that is obtained if treatment is slightly increased for each population member, given the member's existing values for  $T$  and  $\mathbf{X}$ .

## Relationship to the logistic model

- Consider a simple logistic model

$$\Pr(Y_t = 1) = \text{logit}(\alpha + \delta t) \quad \text{where} \quad \text{logit}(z) = \frac{\exp(z)}{1 + \exp(z)}$$

which implies

$$\ln v\{\Pr(Y_t = 1)\} = \alpha + \delta t$$

- Assume  $T$  is binary. We then recover the MOR as

$$\text{MOR} = \exp\{(\alpha + \delta) - (\alpha)\} = \exp(\delta)$$

- Meaning: the (exponent of the) slope coefficient in a simple logistic regression estimates the MOR  
(The same also holds in case of a continuous treatment, which is easy to show.)

## Relationship to the logistic model

- If we condition on  $\mathbf{X}$ , then

$$\ln v\{\Pr(Y_t = 1|\mathbf{X} = \mathbf{x})\} = \alpha + \delta t + \mathbf{x}\beta$$

- Here  $\exp(\delta)$  is the *conditional* odds ratio (i.e., the the odds ratio given a specific value of  $\mathbf{X}$ ).
- The conditional odds ratio (COR) is different from the MOR, which has a more involved form. For example, in case of a binary treatment:

$$\text{MOR} = \exp(\ln v\{E_{\mathbf{X}}[\text{logit}(\alpha + \delta + \mathbf{x}\beta)]\} - \ln v\{E_{\mathbf{X}}[\text{logit}(\alpha + \mathbf{x}\beta)]\})$$

This will be different from COR when  $\beta \neq \mathbf{0}$ .

## Relationship to the logistic model

- The difference between MOR and COR is referred to as *noncollapsibility* or *rescaling bias*.
- “Noncollapsibility of the OR derives from the fact that when the expected probability of outcome is modeled as a nonlinear function of the exposure, the marginal effect cannot be expressed as a weighted average of the conditional effects” (Pang et al. 2016).
- MOR will be attenuated compared to COR (what is commonly referred to as rescaling effects).
- But more importantly:

MOR and COR correspond to different estimands!  
They are conceptually different.



# Why marginal odds ratios?

1. While there exists only one MOR, there are many CORs, as the latter depends on the conditioning set  $\mathbf{X}$ .
2. Given their “on average” interpretation, MORs are easier to compare across different populations and studies (do not depend on arbitrary conditioning sets).
3. MORs behave like AMEs: They can be compared across different conditioning sets and they are “average” effects implied by a model.

- 1 Background
- 2 Marginal (log) odds ratios
- 3 Estimation**
- 4 Example
- 5 Discussion

# Estimation

- Estimand  $\Rightarrow$  Estimation
- There are several approaches how we can estimate the MOR.
  - ▶ G-computation (using predictions from a model)
  - ▶ Inverse probability weighting
  - ▶ Unconditional logistic regression (RIF regression)
- All are discussed in our forthcoming paper (for binary/categorical as well as continuous treatments; including formulas for analytic standard errors based on influence functions).
- Here we focus on G-computation as it closely resembles the formulation of the adjusted MOR above. That is, G-computation obtains the MOR that is *implied* by the chosen logit model. The other methods follow a somewhat different logic.

# G-computation

- G-computation estimates the MOR using counterfactual predictions from a logit model (or any other model in principle).
- For example, for a binary treatment, the procedure is as follows.
  1. Regress  $Y$  on  $T$  and  $\mathbf{X}$  using logistic regression.
  2. Use the model estimates to generate two predictions of  $\Pr(Y = 1)$  for each observation, one with  $T$  set to 0 and one with  $T$  set to 1.
  3. Predictions are then averaged across the sample to obtain estimates of the population-averaged success probability by treatment level.
  4. These average predictions can then be plugged into the formula for the MOR:

$$\ln \widehat{\text{MOR}} = \ln v(\bar{p}^{T=1}) - \ln v(\bar{p}^{T=0})$$

- Note that margins can be used to do the above computations.

# G-computation

- For continuous treatments we evaluate level-specific MORs (using analytic derivatives) at each level of the treatment (possibly using an approximation grid) and then average over the treatment distribution (not directly possible with `margins`).
- An alternative approach is based on applying fractional logit to averaged counterfactual predictions at each value of  $T$ . For binary/categorical treatments this leads to the same results as the procedure above. For continuous treatments results slightly differ (due to the different implicit averaging). Nonetheless we prefer this approach due to its generality and flexibility.

- Software implementing the methods is available from GitHub
  - ▶ <https://github.com/benjann/lnmor>
  - ▶ <https://github.com/benjann/ipwlogit>
  - ▶ <https://github.com/benjann/riflogit>

- 1 Background
- 2 Marginal (log) odds ratios
- 3 Estimation
- 4 Example**
- 5 Discussion

# Example

- Application: gender gap in STEM

```
. use stem, clear
```

```
(Excerpt from TREE cohort 2)
```

```
. describe
```

```
Contains data from stem.dta
```

```
Observations:      6,809
```

```
Variables:         7
```

```
Excerpt from TREE cohort 2
```

```
1 Sep 2022 19:28
```

---

Variable name	Storage type	Display format	Value label	Variable label
stem	byte	%8.0g		Is in STEM training
male	byte	%8.0g		Is male
mathscore	double	%10.0g		Math score
repeat	byte	%8.0g		Ever repeated a grade
books	byte	%19.0g	books	Number of books at home
wt	double	%10.0g		Sampling weight
psu	int	%8.0g		Sampling unit

---

Sorted by:



# Example

- Probability difference

```
. mean stem [pw=wt], over(male) cluster(psu)
```

Mean estimation

Number of obs = 6,809

(Std. err. adjusted for 800 clusters in psu)

	Mean	Robust std. err.	[95% conf. interval]	
c.stem@male				
0	.163234	.0093646	.1448519	.1816161
1	.2748687	.0145161	.2463745	.3033629

```
. regress stem i.male [pw=wt], cluster(psu) noheader  
(sum of wgt is 78,600.1929332293)
```

(Std. err. adjusted for 800 clusters in psu)

stem	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
1.male	.1116347	.0142969	7.81	0.000	.0835708	.1396987
_cons	.163234	.0093653	17.43	0.000	.1448506	.1816174

# Example

- Unadjusted (gross) OR

```
. logit stem i.male [pw=wt], or cluster(psu) nolog
Logistic regression
```

```
Number of obs = 6,809
Wald chi2(1) = 67.37
Prob > chi2 = 0.0000
Pseudo R2 = 0.0172
```

```
Log pseudolikelihood = -40949.278
```

(Std. err. adjusted for 800 clusters in psu)

stem	Odds ratio	Robust std. err.	z	P> z	[95% conf. interval]	
1.male	1.943131	.1572663	8.21	0.000	1.658099	2.27716
_cons	.1950773	.0133746	-23.84	0.000	.1705485	.2231338

Note: \_cons estimates baseline odds.

# Example

- Conventional approach: “conditional” OR

```
. logit stem i.male mathscore i.repeat books [pw=wt], or cluster(psu) nolog
Logistic regression                               Number of obs = 6,809
                                                    Wald chi2(4) = 596.03
                                                    Prob > chi2 = 0.0000
Log pseudolikelihood = -31905.554                 Pseudo R2 = 0.2343
                                                    (Std. err. adjusted for 800 clusters in psu)
```

stem	Odds ratio	Robust std. err.	z	P> z	[95% conf. interval]	
1.male	1.959295	.1675426	7.87	0.000	1.65696	2.316794
mathscore	2.606164	.1252437	19.93	0.000	2.371897	2.86357
1.repeat	.6563627	.0965248	-2.86	0.004	.4920011	.8756321
books	1.087051	.0341241	2.66	0.008	1.022185	1.156034
_cons	.1058314	.0166897	-14.24	0.000	.0776926	.1441616

Note: \_cons estimates baseline odds.

## Example

- G-computation approach (`lnmor` is a post-estimation command, i.e. first estimate the model, then apply `lnmor`)

```
. lnmor i.male, or
```

```
Enumerating predictions:
```

```
male[2]..done
```

```
Marginal odds ratio
```

```
Number of obs = 6,809
```

```
Command = logit
```

```
(Std. err. adjusted for 800 clusters in psu)
```

stem	Odds Ratio	Robust std. err.	t	P> t	[95% conf. interval]	
1.male	1.677032	.1103015	7.86	0.000	1.473911	1.908145

## Example

- Compare results (SEs in parentheses)

---

ln(MOR)	Unadjusted	Conditional	Adjusted
1.male	0.664 (0.0809)	0.673 (0.0855)	0.517 (0.0658)

---

---

MOR	Unadjusted	Conditional	Adjusted
1.male	1.943 (0.157)	1.959 (0.168)	1.677 (0.110)

---

## Example

- Could also do a statistical test for confounding!

```
. quietly logit stem i.male [pw=wt], cluster(psu)
. quietly lnmor i.male, nodots rif(RIF*)
. quietly logit stem i.male mathscore i.repeat books [pw=wt], cluster(psu)
. quietly lnmor i.male, nodots rif(RIFadj*)
. quietly total RIF2 RIFadj2 [pw=wt], cluster(psu)
. lincom RIFadj2 - RIF2
( 1)  - RIF2 + RIFadj2 = 0
```

Total	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
(1)	-.1472752	.0420574	-3.50	0.000	-.2298313	-.0647192

```
. drop RIF*
```

# Some further functionality

- Using `at()` to evaluate interactions

```
. logit stem i.male##c.mathscore##c.mathscore##i.repeat##c.books [pw=wt], ///  
>     cluster(psu)  
     (output omitted)  
. lnmor i.male, nodots or at(repeat)
```

```
Marginal odds ratio                               Number of obs    =    6,809  
                                                    Command        =    logit
```

```
1: repeat = 0
```

```
2: repeat = 1
```

(Std. err. adjusted for 800 clusters in psu)

	stem	Odds Ratio	Robust std. err.	t	P> t	[95% conf. interval]	
1	1.male	1.700763	.1254126	7.20	0.000	1.471573	1.965648
2	1.male	1.514474	.3348553	1.88	0.061	.9812346	2.337496

# Some further functionality

- Using `at()` to evaluate interactions

```
. lnmor i.male, nodots or at(mathscore = -2(2)2)
```

```
Marginal odds ratio          Number of obs    =      6,809  
                             Command                =      logit
```

```
1: mathscore = -2
```

```
2: mathscore = 0
```

```
3: mathscore = 2
```

(Std. err. adjusted for 800 clusters in psu)

	stem	Odds Ratio	Robust std. err.	t	P> t	[95% conf. interval]	
1	1.male	1.697829	.6740845	1.33	0.183	.7788088	3.701323
2	1.male	1.890954	.2009018	6.00	0.000	1.535003	2.329448
3	1.male	1.991302	.3565062	3.85	0.000	1.401245	2.829831



## Some further functionality

- Obtain results for several predictors in one call

```
. lnmor i.male mathscore i.repeat books, or  
(mathscore has 491 levels; using 82 binned levels)
```

```
Enumerating predictions:
```

```
male[2]..mathscore[82].....  
.....repeat[2]..books[7].....done
```

```
Marginal odds ratio          Number of obs    =    6,809  
                             Command              =    logit  
                             (Std. err. adjusted for 800 clusters in psu)
```

stem	Odds Ratio	Robust std. err.	t	P> t	[95% conf. interval]	
1.male	1.676628	.1103018	7.86	0.000	1.47351	1.907746
mathscore	2.647517	.1402908	18.37	0.000	2.385973	2.937731
1.repeat	.772633	.0934318	-2.13	0.033	.6093746	.9796303
books	1.059082	.0245322	2.48	0.013	1.012005	1.108348

## Some further functionality

- Analyze nonlinear effects:

```
. logit stem i.male c.mathscore##c.mathscore i.repeat c.books [pw=wt], ///  
> cluster(psu)  
  (output omitted)  
. lnmod c.mathscore##c.mathscore  
(mathscore has 491 levels; using 82 binned levels)  
Enumerating predictions:  
  mathscore[82] .....  
  .....done  
Marginal log odds ratio                               Number of obs   =    6,809  
                                                       Command         =    logit  
                                                       (Std. err. adjusted for 800 clusters in psu)
```

stem	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
mathscore	1.022883	.0698188	14.65	0.000	.8858334	1.159933
c.mathscore# c.mathscore	-.0757118	.0276166	-2.74	0.006	-.1299215	-.0215022

- 1 Background
- 2 Marginal (log) odds ratios
- 3 Estimation
- 4 Example
- 5 Discussion

# Discussion

- We provide a clear definition of the marginal OR (clarification of estimand).
- We provide flexible software that can estimate the marginal OR for categorical as well as continuous predictors (including support for complex surveys).
- But ...
  - ... is it worth the hassle? How much do applied researchers love odds ratios?
  - ... will it change practice?
- Any other comments/ideas?

# References

- Daniel, R., J. Zhang, and D. Farewell. 2021. Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal* 63(3): 528–557.
- Jann, Ben. 2022. VIOLINPLOT: Stata module to draw violin plots. Available from <https://ideas.repec.org/c/boc/bocode/s459132.html>.
- Karlson, K. B., A. Holm, R. Breen. 2012. Comparing Regression Coefficients Between Models using Logit and Probit: A New Method. *Sociological Methodology* 42: 286–313.
- Karlson, K. B., F. Popham, and A. Holm. 2021. Marginal and Conditional Confounding Using Logits. *Sociological Methods & Research*.
- Pang, M., J. S. Kaufman, and R. W. Platt. 2016. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Statistical Methods in Medical Research* 25(5): 1925–1937.
- Stampf, S., E. Graf, C. Schmoor, and M. Schumacher. 2010. Estimators and confidence intervals for the marginal odds ratio using logistic regression and propensity score stratification. *Statistics in Medicine* 29: 760–769.
- Zhang, Z. 2008. Estimating a Marginal Causal Odds Ratio Subject to Confounding. *Communications in Statistics – Theory and Methods* 38(3): 309–321.

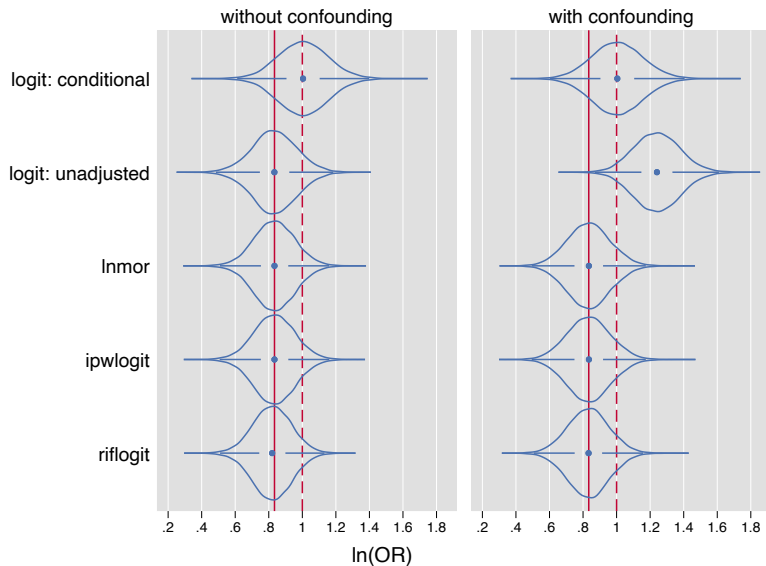
# Appendix: Some simulation results

- Setup

- ▶ Binary outcome  $Y$  depends on treatment  $T$  and control variable  $X$  through a logistic model.
- ▶ The effects of  $T$  and  $X$  on  $Y$  (the conditional log odds ratios) are set to 1 in all simulations (intercept is 0).
- ▶  $X$  has a standard normal distribution.
- ▶  $T$  is either binary or continuous.
- ▶ Two scenarios:
  1. unconfounded:  $T$  is independent from  $X$  ( $X$  has an even distribution if binary and a standard normal distribution if continuous)
  2. confounded:  $T$  depends on  $X$  (binary: logistic model with slope 0.5; continuous: linear model with slope 0.5 and standard normal errors)
- ▶ 10'000 replications.
- ▶ Using `violinplot` (Jann 2022) to display results.

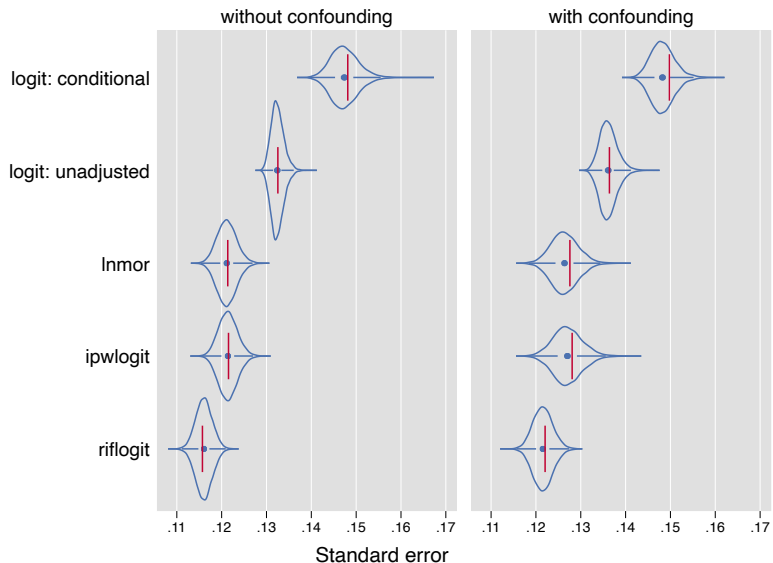
# Appendix: Some simulation results

- Distribution of effect estimates for binary treatment



# Appendix: Some simulation results

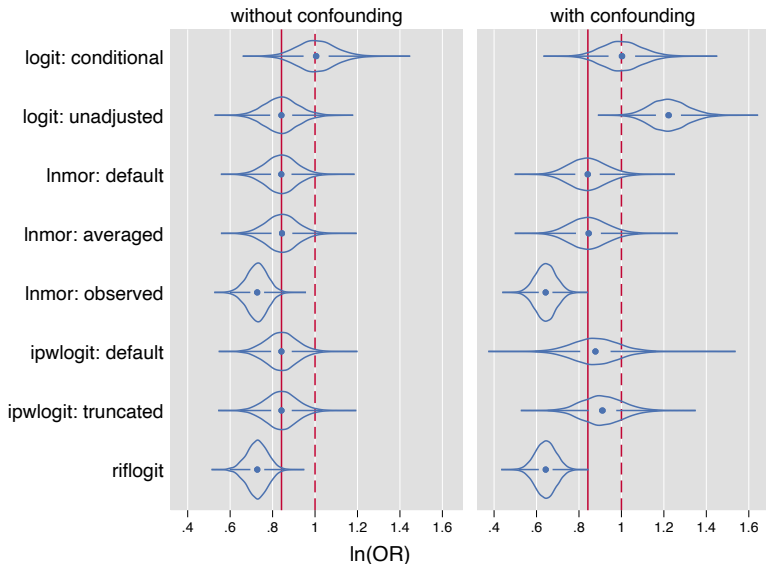
- Distribution of standard errors for binary treatment





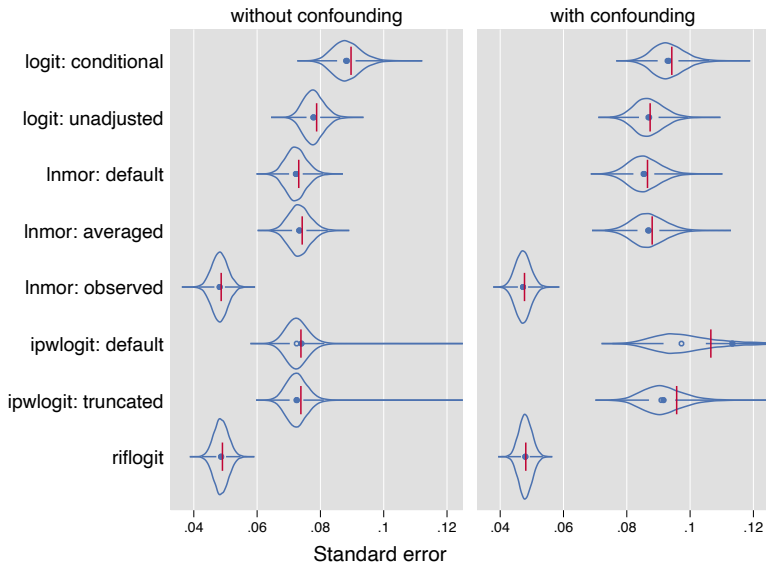
# Appendix: Some simulation results

- Distribution of effect estimates for continuous treatment



# Appendix: Some simulation results

- Distribution of standard errors for continuous treatment



## Appendix: Some simulation results

- Binary treatment:
  - ▶ All estimators appear to work well.
  - ▶ However, note that the treatment has an even distribution in these simulation; may need to do more simulations with uneven distribution.
- Continuous treatment:
  - ▶ IPW does not fully remove confounding. Furthermore, stability of IPW becomes problematic. Truncation helps somewhat but also increases bias.
  - ▶  $MOR'$  ("observed") is a different estimand than  $\overline{MOR}$  ("averaged"). RIF logit appears to approximate  $MOR'$ , not  $\overline{MOR}$ .