Marginal versus conditional effects: does it make a difference?

Mireille Schnitzer, PhD
Université de Montréal

Overview

- In observational and experimental studies, the goal may be to estimate the effect of a binary treatment.
- Researchers often adjust for baseline covariates using logistic regressions. The coefficient of treatment in this model is interpreted as the « effect » (in terms of the log-odds ratio).
- The parameters in a logistic regression are uncollapsible – this means that if you add a covariate that is related to the treatment and/or outcome, the true value of the coefficient of treatment changes.
- My view (which is pioneered by many working in both clinical trials and causal inference) is that it is preferable to try and estimate the marginal odds ratio.

Overview

- In these slides, I present a simple randomized setting where we compare the difference between the true conditional odds ratio and the true marginal odds ratio.

- We see that these parameters (compared on a relative scale) can differ substantially but that this will depend on the **associations** between the variables and the outcome and on the **frequency** of the outcome.

Setting

- Suppose we have baseline data, C, for every subject.

- Binary exposure indicator, A.

- Outcome Y.

- Interest lies in estimating the effect of exposure A on outcome Y.

Traditional effect measures

- In traditional statistical approaches, we propose a model that represents the outcome process, i.e. $E(Y|A,C)$.
  - E.g. A linear/logistic regression

- This model is made conditional on treatment type and all covariates deemed necessary to unconfound the effect estimation (or improve efficiency in a randomized trial).

- Finally, the coefficient of the treatment covariate is interpreted as the effect of interest.

Marginal effect measures

- Each subject has the potential to be exposed to either intervention level (*positivity*).

- Therefore, each subject has two potential outcomes (or *counterfactuals*) Y^1 and Y^0.
  - We only observe the one that corresponds with the intervention actually received.

Adjusted vs conditional

- Marginal effects defined using potential outcomes have a causal interpretation regardless of the estimation methods used.

- The marginal odds ratio, $\frac{P(Y^1=1)/P(Y^1=0)}{P(Y^0=1)/P(Y^0=0)}$, contrasts the odds of obtaining the outcome had everyone in the population received intervention 1 versus 0.

- To estimate marginal effects, it might still be necessary to adjust for confounders.

- Note that there are many available methods to estimate the marginal odds ratio while adjusting for confounders (e.g. inverse propensity score weighting, G-Computation, and Targeted Maximum Likelihood Estimation).

Problems with traditional methods

Non-collapsibility

- If you use a logistic regression model to estimate an effect on a binary outcome, the resulting odds ratio is *conditional* on the covariates in the model.

- This is different from the *marginal* odds ratio, $\frac{P(Y^1=1)/P(Y^1=0)}{P(Y^0=1)/P(Y^0=0)}$.

Problems with traditional methods

Why is this a problem?

- Suppose you have a randomized controlled trial with two arms and a binary outcome. (Full adherence, no dropout, no confounding.)

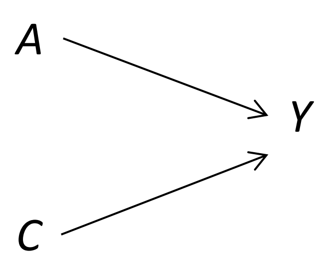
- A logistic regression without covariates gives the *marginal* odds ratio.

- However, if you add outcome predictors to this model, **the parameter of interest changes**.
  - This means that what you mean by « effect » has changed, even though the true effect does not change.
  - This also means that two analysts may choose to adjust for different causes of the outcome and therefore target two different parameter values.

Non-collapsibility without confounding

- Consider a situation where there are no confounders of the *A*-*Y* association.
  - For instance, in an clinical trial where *A* is randomized.
  - *Y* is a binary outcome

- *C* is a pure cause of the outcome.



Comparison between conditional and marginal effect

- We have that Y is generated conditional on A and C:

$$logitP(Y = 1|A, C) = b + b_1 A + b_2 C$$

  - $\exp(b_1)$ is the *conditional odds ratio*

- The *marginal odds ratio* $\exp(b_1^*)$ is defined as:

$$logitP(Y = 1|A) = b^* + b_1^* A$$

$$\exp(b_1^*) = \frac{P(Y^1 = 1)/P(Y^1 = 0)}{P(Y^0 = 1)/P(Y^0 = 0)}$$

Comparison between conditional and marginal effect

- Goal: to compare the *true* values of $\exp(b_1)$ and $\exp(b_1^*)$.
  - This will demonstrate the difference in the *targeted* parameters.
  - With a very large sample size, this represents the estimation bias we would obtain using a conditional logistic regression when the goal is really to estimate the marginal odds ratio.

- We can vary b, b_1, and b_2

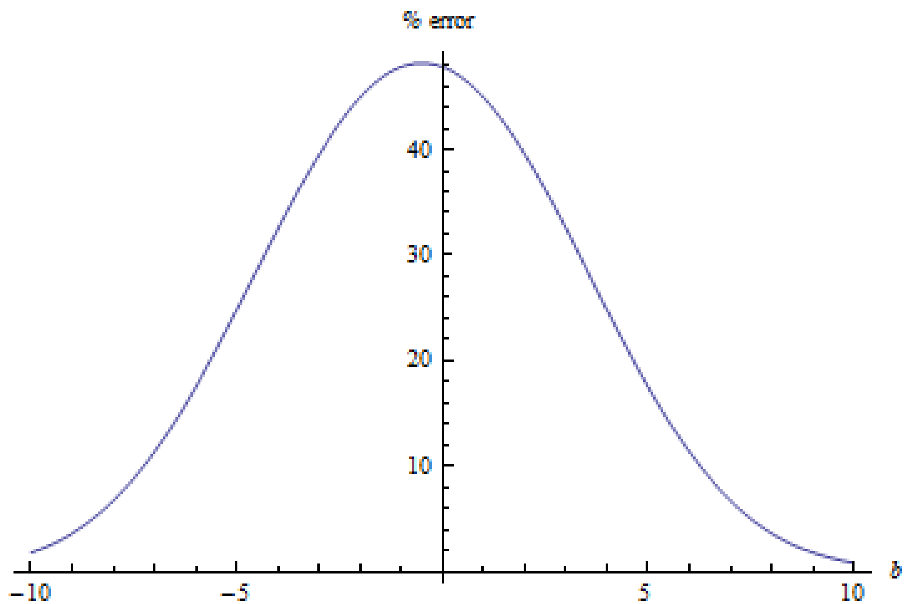
Relative bias measure

- Percent error of the conditional parameter

$$\frac{\exp(b_1) - \exp(b_1^*)}{\exp(b_1^*)} * 100$$

Varying the baseline risk

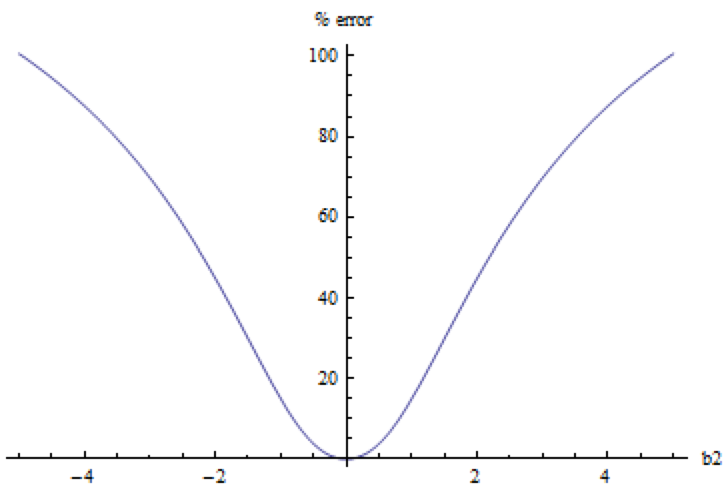
- Let $b_1 = 1$ and $b_2 = 2$.

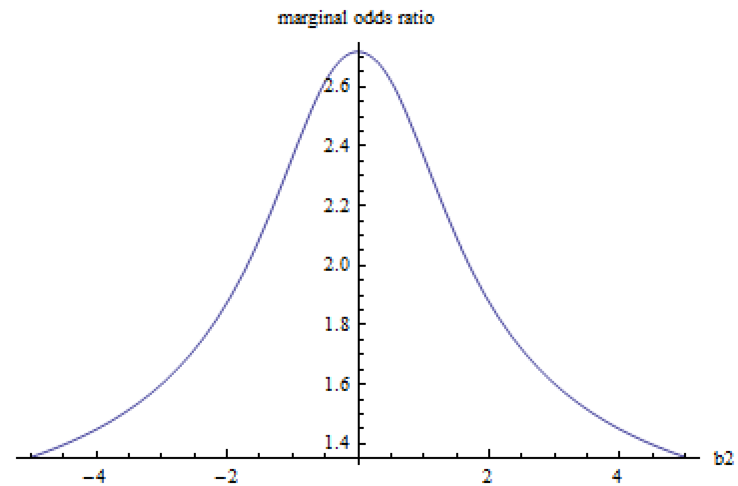


- As the outcome gets extremely rare or common, OR≈RR.
- In the presence of a risk predictor with OR=$\exp(2)$, bias disappears when the baseline *risk* approaches $\exp it(-10)$=0.000045

Varying the strength of the risk predictor

- Let $b_1 = 1$ and $b = 0$

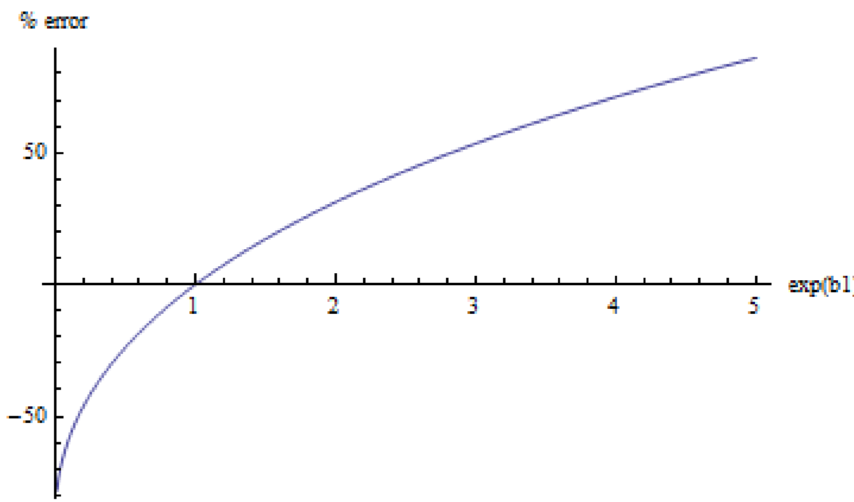




- The marginal odds ratio decreases when the strength of the risk factor increases.
- The % error increases as the strength of the risk predictor increases (no bias when it is not a predictor).

Varying the conditional odds ratio

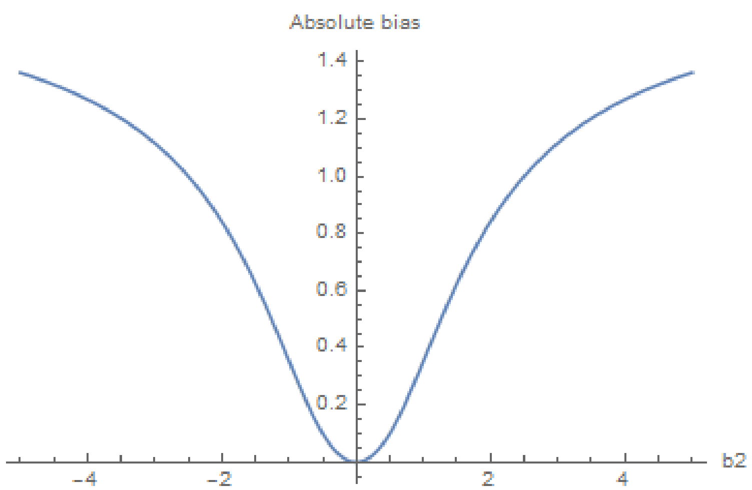
- Let $b = 0$ and $b_2 = 2$

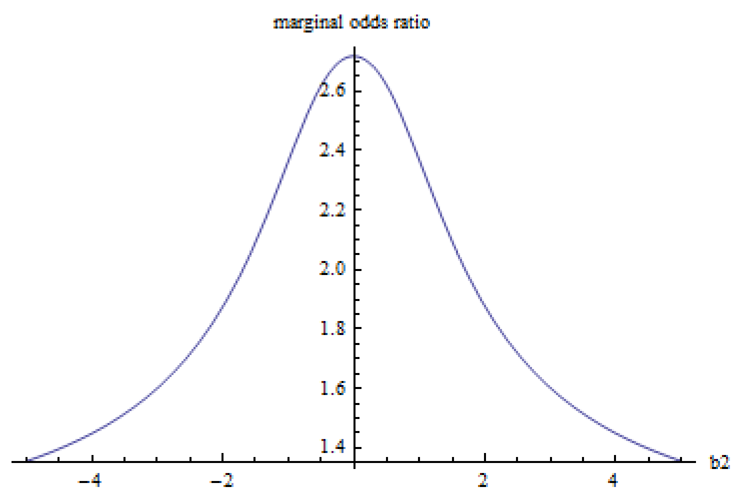


- For negative effects, the bias is *negative*.
- For positive effects, the bias is *positive.*
- Regardless of direction of risk predictor, the bias is always *anti-conservative*.

Absolute bias when varying the strength predictor

- To see the magnitude of the bias, define the absolute bias as $\exp(b_1) - \exp(b_1^*)$.
- Let $b_1 = 1$ and $b = 0$





Additional issues

- Because using a logistic regression for the estimation of the conditional odds ratio assumes a full parametric structure, confidence intervals will be narrow.
  - In general parametric assumptions are likely flawed.
  - So not only will the conditional odds ratio always inflate the apparent magnitude of the association, but it will also be « overly confident » about it.

- In the numerical examples, we assumed that the conditional logistic regression was correct.
  - In practice, it is not, potentially leading to additional bias.

But what if I want the conditional OR?

- If you are interested in effect modification, you might reasonably fit

$$logitP(Y = 1|A, C) = b + b_1 A + b_2 C + b_3 AC$$

in order to investigate b_3.

- However, I argue that all real-world interpretations of b_1 and b_1^* are indifferentiable.

- I.e. if you have only main-terms in the model, you are essentialy interpreting b_1 as b_1^*. And this can be entirely misleading.

Do your own investigation

- I provide the Mathematica code that I used to compute the true values of the marginal and conditional odds ratios and make the plots.

- Change the parameter values (of b, b_1, and b_2) and see how this affects the results.

- You can also investigate this issue using simulated datasets with large sample sizes.