Risk ratio regression - simple concept and simple computation

Frank Popham (frank.popham@protonmail.com)

2023-02-07

Dear Editors,

A new IJE paper states in its title that "Risk ratio regression - simple concept yet complex computation"[1]. This is only true if one wants to read the risk ratio directly from the coefficients of your model. Given a binary outcome and binary exposure as in the aforementioned paper, a logistic regression is the "natural" choice. While its coefficients will be (log) odds ratios, it is simple to derive a number of other effect measures including the risk ratio. This can be done easily using modern software such as R (see accompanying code).

In the paper under discussion the risk of weight gain relative to quitting smoking or not was studied. Using standardization (g formula)[2], I easily estimate a risk ratio . The method is simple,

1) fit the model of outcome by exposure and confounders using a logistic regression model.
2) from this model predict for each person the probability of the outcome treating everyone as exposed (E) and then everyone as not exposed (NE) (everyone quit or no-one quit in our example).
3) Average these probabilities for each of the two scenarios. We can then compare these two average predictions to obtain an absolute difference (E-NE), the risk ratio (E/NE), or the odds ratio (E/(1-E)) / (NE/(1-NE)). See Table 1.

The first stage retains the advantages of a logistic model for a binary exposure in that the model usually converges and predicted probabilities will be in the range of 0 to 1. The second and third stage avoid non-collapsibility as we predict probabilities (collapsible) rather than odds (non-collapsible) before averaging across the strata from the stage 1 model.

Table 1 - Losing weight by quitting smoking

	Quit smoking	Estimate	95% CI - low	95% CI - high
Absolute	No	46.4%	43.5%	49.2%
Absolute	Yes	60.7%	55.9%	65.5%
Difference	Yes-No	14.3%	8.7%	20.0%
Risk ratio	Yes/No	1.31	1.18	1.45
Odds ratio	(Yes/(100%-Yes)) / (No/(100%-No))	1.79	1.42	2.26

It should be noted that the odds ratio from the stage 1 model (1.84) is not the same as in Table 1 as the former is a conditional odds ratio while the latter (and all effects in Table 1) are marginal. We can use standardization to obtain the stage 1 model odds ratio by predicting the log odds at stage 2 rather than the probability.

In conclusion the risk ratio is easily obtainable from a logistic regression. Being clear about whether we are reporting marginal and conditional estimates is another important consideration and authors should be explicit about the effect measure reported.

Best wishes,

Frank Popham

References

1. Mittinty MN, Lynch J. Reflection on modern methods: risk ratio regression - simple concept yet complex computation. *International Journal of Epidemiology* [Internet]. 2022 Nov 23; Available from: http://dx.doi.org/10.1093/ije/dyac220
2. Hernan MA, Robins JM. Causal inference: What if. CRC Press; 2020;