Tidsskriftet Tidsskrift for Den norske legeforening

Logistic regression with more than two categories

MEDICINE AND NUMBERS

STIAN LYDERSEN

stian.lydersen@ntnu.no
Stian Lydersen, PhD, professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Norwegian University of Science and Technology.
The author has completed the ICMJE form and declares no conflicts of interest.

When the outcome variable has only two levels or categories, standard binary logistic regression can be used. When it has three or more, we can use other variants of logistic regression.

Let us start with an example: Munthe-Kaas et al. (1) studied a possible association between frailty before a stroke and degree of cognitive impairment three months after a stroke in 598 patients. The independent variable, frailty, was measured on an index that in theory ranges from 0 to 1, where a higher value indicates greater frailty. Before the stroke, the patients scored from 0 to 0.56, with a mean value of 0.14 and a standard deviation of 0.10. The dependent variable, degree of cognitive impairment, is ordinal with three categories. Three months after the stroke, 286 (45 %) patients had normal cognition, 172 (29 %) had mild cognitive impairment, and 158 (26 %) had dementia.

Separate binary regression models

When the dependent variable is ordinal, such as here, we can choose to dichotomise it and use binary logistic regression. A variable with three categories can be dichotomised in two ways (Figure 1). If we set the threshold

between dementia and mild cognitive impairment, we will have the category of dementia on one side and the combined category of mild impairment and normal cognition on the other. The appurtenant odds ratio per 0.10 units of increase in frailty will be 3.09 (95 % confidence interval (CI) 2.45 to 3.89, $p < 0.001$). If we set the threshold between normal cognition and mild cognitive impairment, the odds ratio will be 2.29 (CI 1.83 to 2.87, $p < 0.001$).

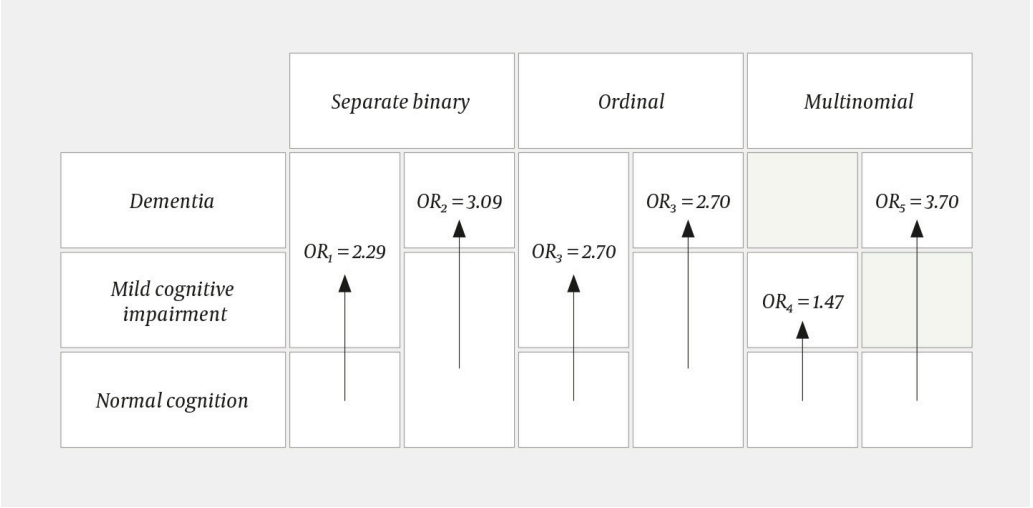

Figure 1 Alternative logistic regression models with odds ratio (OR) per 0.10 units of increase in the frailty index.

Ordinal logistic regression

Alternatively, ordinal logistic regression can be used. It comes in several versions (2), and the one most frequently used is called 'proportional odds logistic regression'. As above, we then perform an analysis for every possible threshold between the categories in the dependent variable, but now with the simplified assumption that the odds ratio is the same for each threshold. In our example, we then obtain an odds ratio of 2.70 (CI 2.23 to 3.27, $p < 0.001$).

The assumption that the odds ratio is the same for each threshold is called the 'proportional odds assumption'. A hypothesis test for this assumption in our example yields a *p*-value of 0.014. This indicates that the assumption is not fully met. Ordinal logistic regression may nevertheless be suited to examine a research question (3, p. 315–6). A hypothesis test of this assumption may thus be of limited relevance. In their study, Munthe-Kaas et al. (1) chose separate binary regression models, not only based on this hypothesis test, but in particular to investigate whether frailty was a stronger predictor for the distinction between dementia and mild cognitive impairment than for the distinction between mild cognitive impairment and normal cognition.

Multinomial logistic regression

If the categories in the dependent variables are not ordinal, we can use multinomial logistic regression. We will then need to choose a reference category. In our example, it is natural to choose normal cognition as the

reference. We then obtain an odds ratio for mild cognitive impairment of 1.47 (CI 1.12 to 1.92, $p = 0.06$) and an odds ratio for dementia of 3.70 (CI 2.81 to 4.87, $p < 0.001$), both of which relative to normal cognition. Note that the odds ratio for mild cognitive impairment of 1.47 in this case is lower that odds ratio of 2.29 obtained by binary logistic regression. The reason is that in the binary logistic regression we are looking at the odds ratio for the combined categories of mild cognitive impairment and dementia, while in multinomial logistic regression we are looking at mild cognitive impairment alone. Correspondingly we see that the odds ratio of 3.70 obtained by multinomial logistic regression applies to normal cognition, while the odds ratio of 3.09 obtained from binary logistic regression is lower because it applies to mild cognitive impairment and normal cognition combined.

Choice of model

If the dependent variable is not ordinal, a multinomial logistic model will be the most natural choice. If the dependent variable is ordinal, we can choose ordinal logistic regression if the proportional odds assumption gives a good approximation to the data or if it is of no practical interest to distinguish between effects at the different levels of the dependent variable. Compared to separate binary regression models, one of the advantages of ordinal logistic regression is that it includes fewer unknown quantities, here odds ratios, in the model. This results in a considerable simplification of the model, especially if there are more than three categories in the dependent variable or if there are multiple predictors.

The three alternative regression models described here have the same interpretation of odds ratio as in binary logistic regression: the odds ratio is the relative increase in odds for each unit of increase in the predictor.

REFERENCES

1. Munthe-Kaas R, Aam S, Saltvedt I et al. Is Frailty Index a better predictor than pre-stroke modified Rankin Scale for neurocognitive outcomes 3-months post-stroke? BMC Geriatr 2022; 22: 139. [PubMed][CrossRef]

2. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 3. utg. Hoboken, NJ: Wiley, 2013.

3. Harrell FE. Regression modeling strategies. With applications to linear models, logistic and ordinal regression, and survival analysis. 2. utg. Cham: Springer, 2015.

Publisert: 23 June 2022. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0786