Injury Prevention 1997; **3:** 294

Logistic regression and odds ratios

Robert W Platt

McGill University/ Montreal Children's Hospital Research Institute, 2300 Tupper, Montreal, PQ H3H 1P3, Canada

Correspondence to: Dr Platt.

Epidemiologists and public health specialists are often interested in assessing outcomes that are binary, that is they take on two mutually exclusive values such as yes or no, dead or alive, drowned or not drowned. We are usually interested in associating such outcomes with several predictors and examining the potential confounding effect of certain predictors. ('Confounding' is the effect of a third variable on the direction or magnitude of an association between an exposure and outcome.) Analyses to assess such outcomes are usually best done using logistic regression, a convenient method for associating a binary outcome with one or more confounders. In this column, I address the structure of logistic regression models and the inferences that can be made from these models.

Logistic regression is an extension of linear regression. Linear regression associates a response, or dependent variable, or outcome y, to an independent or predictor variable (or variables) x. Linear regression gives predictions of the outcome through models that take the general form $E(y)=\beta_0+\beta_1 x$. In this equation $E(y)$ represents the mean or average value of the outcome. In contrast, when logistic regression is used, the outcome is no longer a continuous variable but rather the probability that it will take on one of the two values, and it is this probability that we estimate.

Consider an example where we want to associate drowning to various parental and environmental factors. We would then be interested in predicting the probability of drowning as it relates to these factors. Linear regression could be used in some situations to do so, but it can lead to serious problems and is difficult to interpret. A mathematical rearrangement of the model would be needed so that we can have a model in a useful form, and this would make the β parameters in logistic regression mean something different than those in ordinary linear regression. In ordinary regression, β_1 refers to the average change in y for a unit change in x. In logistic regression, $e^{\beta 1}$ refers to the odds ratio of the outcome (for example drowning) for two subjects that differ by one unit of x (for example parental supervision). The odds ratio is similar to, but somewhat different from, the more commonly understood relative risk or risk ratio. The relative risk for comparing the outcome in one group to that in another group is simply

$$RR = \frac{Risk_1}{Risk_2},$$

the ratio of the risks, or probabilities of the outcome, in the two groups. The odds ratio is slightly different,

$$OR = \frac{Odds_1}{Odds_2},$$

where the odds of the disease in each group is given by

$$Odds_i = \frac{Risk_i}{1-Risk_i}.$$

For example, a risk of disease of 0.5 (50%) corresponds to an odds of disease of 1.0, and a risk of 0.2 corresponds to an odds of 0.2/0.8=0.25.

Usually, the relative risk is what we would like to know, so why do we use the odds ratio? One reason is mathematical convenience—logistic regression easily produces estimates of the odds ratio. Further, in case-control studies, the odds ratio is the only parameter comparing the two groups that we can estimate. (It is, however, comforting to learn that when the disease or outcome being analyzed is 'rare', the odds ratio is nearly the same as the relative risk.) This happens because when the outcome is rare, $1-Risk_i$ is approximately 1. For example, if the risk in one group is 0.1% and in the other 0.2%, the risk ratio or relative risk is 2; in the same situation the odds ratio is

$$^{0.02}/_{0.98}/^{0.01}/_{0.99} = 2.02.$$

The two results are virtually indistinguishable.

Logistic regression is easy to do using most popular statistical software packages like SPSS or SAS, and the results are generally readily interpretable. It is important to remember what the parameters are, though, and what they mean. A basic introduction to this topic can be found in the workbook text by Kleinbaum.[1]

1 Kleinbaum DG. *Logistic regression: a self-learning text.* New York: Springer-Verlag, 1984.

Further reading

Hosmer DW, Lemeshow S. *Applied logistic regression.* New York: John Wiley, 1989.