READERS' OPINION AND DISCUSSION

Can studies where subjects have different follow-up times be analysed through binomial regression?

Neal Alexander/+

Unidad de Epidemiología y Bioestadística,
Centro Internacional de Entrenamiento e Investigaciones Médicas,
Cali, Colômbia

Penna (2011) asks whether subjects with different follow-up times can be analysed through binomial regression. The answer to this question is "yes".

If the rate for a particular person is λ and they have been observed for a period of time (t), then the probability (p) of having an event during that time is equal to $1\text{-}e^{-\lambda t}$. The equation can be re-arranged to give $\log[-\log(1\text{-}p)] = \log(\lambda)+\log(t)$. This is a linear function of the log rate, enabling it to be modelled by regression, while the follow-up time values (t) can vary between individuals (Collett 1991).

This approach is useful when the exact time of the event is unknown as, for example, in seroconversion studies like that of Gray et al. (2001). The regression technique is a generalised linear model with "complementary log-log link function"; the logarithm of time is used as an "offset". As it uses the binomial distribution family, the heading "binomial regression" is appropriate. However, it differs from the model used by Mastrangelo et al. (2011) in that it yields rate ratios rather than risk differences.

REFERENCES

Collett D 1991. *Modelling binary data*, Chapman and Hall, London, 369 pp.

Gray RH, Wawer MJ, Brookmeyer R, Sewankambo NK, Serwadda D, Wabwire-Mangen F, Lutalo T, Li X, van Cott T, Quinn TC 2001. Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet 357*: 1149-1153.

Mastrangelo G, da Silva Neto J, da Silva GV, Scoizzato L, Fadda E, Dallapicola M, Folleto AL, Cegolon L 2011. Leprosy reactions: the effect of gender and household contacts. *Mem Inst Oswaldo Cruz 106*: 92-96.

Penna MLF 2011. Can studies where subjects have different follow up times be analyzed through binomial regression? *Mem Inst Oswaldo Cruz 106*: 383-384.

+ Corresponding author: nalexander@cideim.org.co
Received 31 May 2011
Accepted 19 July 2011

REPLY

Comments in regards to Penna's and Alexander's letters

Penna (2011) and Alexander (2011, this letter) discuss the use of logistic regression for the analysis of incidence data. The motivation is taken from Mastrangelo et al. (2011), who used logistic regression to assess the effect of gender and household contacts as they relate to the risk of leprosy reactions. The study population was comprised of leprosy patients in treatment from 1998-2005. In spite of availability of the date of leprosy reaction for each patient, the authors used logistic regression and treated the problem as prevalence instead of incidence. Penna (2011) emphasises the problem that different follow-up times should not be ignored. Alexander argues that it is possible to incorporate different follow-up times in a binomial regression using a log-log link function with the inclusion of follow-up time for each subject in the model. This is a useful approach when the exact time of the event is unknown. This method for data analysis was not conducted by Mastrangelo et al. (2011). Both Penna (2011) and Alexander (2011) are correct in their comments, which contribute to a better understanding as to the appropriate use of statistical models and the best possible use of available data.

Marilia Sá Carvalho
Claudia Torres Codeço
Programa de Computação Científica - Fundação Oswaldo Cruz