Logistic regression in case-control studies: the effect of using independent as dependent variables.

Nagelkerke, NJ; Moses, S; Plummer, FA; Brunham, RC; Fish, D

Abstract:

In case-control studies, cases are sampled separately from controls. In such studies the primary analysis concerns the estimation of the effect of covariables on being a case or a control. To explore causal pathways, further secondary analysis could concern the relationships among the covariables. In this paper the validity of such secondary analysis is addressed. In particular, the use of multiple logistic regression in case-control studies where the dependent variable is not the case/control indicator is explored. It is shown that only under very restrictive conditions will sample regression coefficients correctly estimate their true value. In many situations, it may be valid to regress one covariable on others in the control group, but not in the case group or the combined sample. This principle is illustrated by a study of sexually transmitted disease in Kenya.