Statistical interactions and Bayes estimation of log odds in case-control studies

Jaya M. Satagopan*
Memorial Sloan Kettering Cancer Center New York, NY, USA– satagopj@mskcc

Sara H. Olson
Memorial Sloan Kettering Cancer Center, New York, NY, USA olsons@mskcc.org

Robert C. Elston,
Case Western Reserve University, Cleveland, OH, USA robert.elston@cwru.edu

We focus on the estimation of the logarithm of disease odds (log odds) when evaluating two risk factors, whether or not interactions are present. The log odds is an important epidemiologic parameter that can facilitate the estimation of odds ratios and absolute risk of disease. It can also provide insights into the potential benefits of screening high-risk individuals. Hence, it is crucial to obtain accurate and precise estimation of log odds in practical settings. We typically fit a logistic regression model to case-control data and estimate the log odds using the fitted model. When evaluating two (or more) risk factors in relation to disease risk using a logistic regression model, interaction terms may be required in order to obtain a good fit to the data. Statisticians define interaction as a departure from an additive model on a certain scale of measurement of the outcome. Certain interactions, known as removable interactions, may be eliminated by fitting an additive model under an invertible transformation of the outcome. For binary disease traits, a transformation is a link function. We show that, when there is an interaction between two risk factors on the logistic scale and when this interaction is removable: (i) the Guerrero and Johnson (GJ) link function is a suitable transformation to additivity; and (ii) fitting an additive model to the data under the GJ link can provide accurate and more precise estimates of the log odds. In practice, we may also encounter non-removable interactions. The model must then include interaction terms, regardless of the choice of the link function. However, in practical settings, we do not know at the outset whether an interaction exists, and if so whether it is removable or non-removable. Rather than trying to decide on significance levels to test for the existence of removable and non-removable interactions, we develop a Bayes estimator based on a squared error loss function. We demonstrate the favorable bias-variance trade-offs of our approach using simulations, and provide empirical illustrations using data from three published endometrial cancer casecontrol studies.

Keywords: Guerrero-Johnson link, Logistic link, Mean squared error, Minimax estimator.