HIERARCHICAL LOGISTIC REGRESSION IN A MULTICENTRIC STUDY OF MULTIPLE DIETARY EFFECTS ON A DISEASE OUTCOME: A FULLY BAYESIAN APPROACH

G. Roli[†1], P. Ferrari[2], E. Riboli[2], P. Monari[1], G. Cavrini[1]

[1]*University of Bologna, Bologna, Italy;*
[2]*International Agency of Research on Cancer, Lyon, France*

[†] E-mail: *roli@stat.unibo.it*

A hierarchical logistic regression model is considered for nested observations, in order to improve relative risk estimates of dietary factors on colon-rectum cancer. The data consist of a sample of 19,693 subjects nested into 27 recruitment centers from the EPIC (European Prospective Investigation into Cancer) study. The hierarchical regression model takes into account the variability associated to different levels of nesting (i.e., individuals within EPIC centers), and the consequent dependence between observations. Moreover, in a second-stage model, prior information on food effects is considered, thus allowing the correlations among dietary effects to be modelled. The prior information refers to the nutrient contents of each food. As a consequence, in this way the final estimates are pulled toward each other when the corresponding variables have similar levels of nutrients (exchangeability assumption). For this analysis a total of 30 foods and 14 nutrients were considered. Fully Bayesian estimates are computed, since Bayesian perspective provides a flexible and complete framework for hierarchical analysis and prior assumptions. When compared with one-stage maximum-likelihood logistic regression, our Bayesian hierarchical regression gives more stable estimates. In particular, large ordinary estimates become more reasonable and less biased.