Clustering on factor analysis dietary patterns: the risk of breast and ovarian cancers

Valeria Edefonti[1], Adriano Decarli[1,2], Giorgia Randi[1,3] and Monica Ferraroni[4]

[1] Istituto di Statistica Medica e Biometria "G. A. Maccacaro", Università degli Studi di Milano, ITALY
[2] Fondazione IRCSS Istituto Nazionale dei Tumori, Milano, ITALY
[3] Dipartimento di Epidemiologia, Istituto di Ricerche Farmacologiche "Mario Negri", ITALY
[4] Dipartimento di Medicina, Chirurgia e Odontoiatria, Università degli Studi di Milano, ITALY

Dietary issues in cancer have been traditionally considered in terms of single foods and nutrients. Because of the complexity of diet and the potential for interaction between food components, these approaches may miss information on the role of diet in cancer etioiology. More recently, nutritional epidemiologists have introduced dietary patterns, intended as combinations of foods, nutrients, or foods and nutrients used to represent the total diet or key factors of the diet. Dietary patterns have been used to describe associations between diet and disease and to characterize a healthy diet in an actual population, allowing for dissemination of dietary recommendations in a more practical way [1]. The so called "a posteriori" dietary patterns are generally defined referring to multivariate statistical methods like factor and cluster analyses. In the following paper we combined them in an overall strategy for data reduction and clustering and applied the procedure to a multicentric case-control study on breast and ovarian cancers conducted in Italy between 1991 and 1999 and based on a validated food-frequency questionnaire [2].
An exploratory principal component factor analysis was performed on the correlation matrix of a selected set of 30 major macro and micronutrients [3]. After varimax rotation, we retained four factors, which explained 75.7% of the total variance in the original dataset. Dietary patterns were obtained from a K-means clustering procedure on factor scores from this factor analysis. To find the most reasonable number of clusters, we ran a series of analyses with predefined cluster numbers and compared a set of statistics built on several indices between all runs. We chose the final solution after having scrutinized the food-intake patterns of the first two ones. Odds ratios (OR) for both cancers were estimated using unconditional multiple logistic regression models on clusters of patients and suitable confounding variables. Floating absolute risk method was used for reporting 95% floating confidence intervals (FCI) [4].
We identified five groups of subjects. The G3 cluster, including subjects more likely to have the lowest intakes of any food group, was used as reference. Belonging to the G5 cluster, including subjects more likely to consume bread and pasta, was unfavorable for both cancers. Conversely, belonging to the G1 group, including subjects that were more likely to consume fruits and vegetables, was protective against ovarian cancer.

References

[1] Newby PK, Tucker KL (2004) Empirically derived eating patterns using factor or cluster analysis: a review. Nutr Rev 62(5):177-203

[2] Franceschi S, Favero A, Decarli A, et al. (1996) Intake of macronutrients and risk of breast cancer. Lancet 347(9012):1351-1356

[3] Edefonti V, Decarli A, La Vecchia C, et al. (2008) Nutrient dietary patterns and the risk of breast and ovarian cancers. Int J Cancer 122(3):609-613

[4] Plummer M (2004) Improved estimates of floating absolute risk. Stat Med 23:93-104