

## Special Report

# Dietary patterns: report of an international workshop

K Hoffmann<sup>1,\*</sup>, MB Schulze<sup>1</sup>, H Boeing<sup>1</sup> and HP Altenburg<sup>2</sup>

<sup>1</sup>Department of Epidemiology, German Institute of Human Nutrition, Arthur-Scheunert Allee 114–116, 14558 Bergholz-Rehbrücke, Germany: <sup>2</sup>Division of Clinical Epidemiology, German Cancer Research Center, Heidelberg, Germany

An international workshop on dietary patterns was held in Potsdam, Germany, on February 22–23, 2000. The workshop was attended by 28 participants from 7 countries. It was organised by the Department of Epidemiology of the German Institute of Human Nutrition. The aim of the workshop was to exchange experiences in determining dietary patterns and to discuss the application of various multivariate statistical methods.

Dietary patterns are of considerable interest for nutritional epidemiology to reflect the complexity of dietary intake in relation to diseases. Traditional analyses based on single nutrients and foods sometimes fail since interactions among them cannot be taken into account. Moreover, ordinary and logistic regression models with many correlated exogenous variables are rather unstable and liable to large confidence intervals for the regression parameters. Therefore, in pursuit of a more meaningful and stable approach, summary variables reflecting dietary intake should be determined. Multivariate statistical methods such as factor analysis or cluster analysis might be helpful. Both methods can be considered complementary to each other. Cluster analysis is a subject oriented method with the objective of joining together different subjects to subgroups, whereas factor analysis is oriented to variables with the aim of reducing the numbers of explaining variables. Two more complex approaches in this field are the latent variable analysis and the structural equation modelling.

E. Wirfält (Malmö, Sweden) started the first day of lectures by describing how to apply cluster analysis to examine food patterns of the total diet. The results presented were based on a data set from baseline examinations of the Malmö Diet and Cancer study, between 1991 and 1996. Clustering variables were defined as the percent contribution of specific food groups to total energy intake leading to a six cluster solution producing conceptually well separated food patterns. A further standardisation of the percent energy food group variables using z-scores produced dissimilar results indicating the necessity to select the variable type in accordance with the specific purpose and research

objective of the study. In the second main lecture of the first day F. Hu (Boston, USA) presented a successful application of exploratory factor analysis to identify major dietary patterns. The two predominant patterns, called 'Prudent' and 'Western', showed a good reproducibility since they were qualitatively similar across two food frequency questionnaires and diet records. Moreover, when using the data set of the Nurses' Health Study, it was shown that high Western diet scores were associated with a high risk of coronary heart disease.

These presentations initiated many discussions concerning the impact of measurement errors of the obtained clusters and factors. This discussion was intensified by a presentation by K. Michels (Cambridge, UK) about misspecifications of dietary patterns assessed with different dietary instruments in the EPIC-Norfolk study. H. Boeing (Potsdam-Rehbrücke, Germany) formulated general requirements on dietary patterns including validity and reproducibility.

In the second part of the workshop, leading German specialists on multivariate statistics and its application in social sciences presented an overview on confirmatory factor analysis and structural equation modelling. F. Faulbaum (University of Duisburg, Germany) explained basic concepts, the application and interpretation of several tests and the way to improve a given model. J. Reinecke (University of Münster, Germany) gave some examples and compared the most important software packages available in this field. The participants agreed that confirmatory factor analysis is a welcome statistical tool to examine hypothetically driven patterns, as well as patterns explored within other data sets. Also, structural equation modelling was recognised as a useful method to reflect complex relations between nutrition and continuous variables describing diseases and symptoms. Unfortunately, structural equation models are not suitable to involve dichotomous endogenous variables since they are based on ordinary regression models. Thus, methods such as logistic regression, as one of the most applied statistical model procedures in epidemiology, cannot be involved in the complex structural equations model context. Furthermore, an overview was given on software

available for factor analysis, latent class analysis and structural equation modelling.

Sessions geared towards controversial issues aimed at clarifying remaining important questions and uncertainties. Such questions concerned the homogeneity of the population, the choice of a narrow number of food groups in advance, the use of percent energy food group variables instead of servings, the standardisation or normalisation of variables, the possibility of biologically driven patterns, methods to find a conceptual chain of determinants of patterns, the advantages and disadvantages of cluster and factor analyses, the validity and application of these statistical methods, the use of 'add up' food scores or hypothetically driven indices, and the consideration of measurement errors in pattern analysis and how to reduce them. Final answers to most of these controversies will be subjects for discussion in future meetings.

Further activities of workshop participants were initiated in order to enlarge the empirical database on specific aspects of multivariate food pattern analysis and error structure models. The rationale of these approaches

were mainly to summarise the experiences of different groups through collaborative effort. Three actions are currently under preparation:

to analyse a common data set with nutritional and confounder variables for dietary pattern and their relation to blood pressure

to analyse whether predefined dietary pattern exist in data sets from the US and Europe using confirmatory factor analysis

to identify the error structure in a simulated data set.

Altogether, there will be progress in applying multivariate statistical methods in dietary pattern analysis in the near future. The workshop demonstrated that complex statistical models can compress existing information of many dietary variables to some key variables and are in principle suitable to describe structural aspects of dietary intake of individuals. However, many basic questions still prevail and need to be solved in the future. Such a discussion among interested scientists might also help to provide wider acceptance of these methods among epidemiologists.