Andrea Wendt
Dr. sc. hum.

Dietary factors, dietary patterns and the occurrence of chronic diseases in the EPIC-Heidelberg cohort study

DKFZ (Deutsches Krebsforschungszentrum)
Doktorvater: Prof. Rudolf Kaaks

It has since long been hypothesized based on epidemiologic studies that particular foods or food components are specific causes of the development of type 2 mellitus (T2DM), cardiovascular diseases and cancer. However, the difficulty in epidemiologic studies to disentangle the associations of foods or food components with disease from confounding effects complicates the interpretation of diet-disease associations. Thus, it has been suggested to evaluate diet in association with disease on a higher structural level of empirically derived dietary pattern variables. A particular method that has been often used for this purpose is factor analysis, which aims to express a larger set of interrelated variables in terms of a drastically reduced set of behavioural dimensions. The aim of the present thesis was to evaluate whether a factor analysis can indeed be used for a clear description of food consumption habits in terms of certain behavioural patterns of food use and whether only a few factors can be used to interpret diet-disease associations and to efficiently predict diet-related disease risks in terms of a reduced set of behavioural dimensions of food use.

In the EPIC-Heidelberg cohort study including 21,554 subjects, the correlations between 37 food groups were rather weak, with highest correlations of 0.71 for raw vegetables and plant oils, 0.39 for cooked vegetables and sauce and -0.38 for whole-grain bread and refined grain bread. Based on the weak correlations, the explained variance in subjects' reported food intake levels increased slowly with increasing numbers of extracted factors (more precisely principal components; the principal component method of factor analysis is commonly used in nutritional epidemiology for the definition of dietary pattern variables; subsequently extracted principal components explain a maximum possible amount of variance in original data). In parallel, the ability of increasing numbers of factors to predict diet-related risks of myocardial infarction (MI) and T2DM increased only gradually, too. In comparison, a conventional automatic selection of foods for risk prediction showed a steeper increase in

predictive ability, and for any number of selected foods the predictive ability was higher than for the same number of subsequent factors.

In accordance to the usual procedure for the definition of dietary pattern variables, three successive principal components that appeared previous to an apparent break in the variance diagram (or, alternatively, scree plot) were rotated with the *Varimax* method so as to optimize them in terms of loadings for foods (or, alternatively, correlations with foods) as close to one as possible for a few foods and as close to zero as possible for all other foods. The three factors had comparably high loadings of $\geq|0.3|$ for the following foods: *factor 1*: whole-grain bread, breakfast cereals, vegetarian dishes, fresh fruits, cheese and milk and milk products (positive loadings), and red meat, processed meat, refined-grain bread, fried potatoes and sauce (negative loadings); *factor 2*: raw vegetables, cooked vegetables, plant oils, fresh fruits and sauce (positive loadings); *factor 3*: sweet bread spreads, confectionery, cakes and desserts (positive correlations), and wine and beer (negative loadings).

The pattern of inverse associations of the first factor with MI, T2DM and smoking-related cancer actually mirrored the patterns of inverse and positive associations of its high positive and high negative loading foods, respectively, with risks of MI, stroke, T2DM and/or smoking-related cancer. However, other foods that were inversely or positively associated with disease risks had high loadings on the second or third factor that overall showed associations with disease risks different from the particular foods or they did not load high on any of the extracted factors. In this respect, a conventional analysis of diet-disease associations is actually required in order to identify food intake habits in subjects with low or high disease incidence.

Indeed, a conventional analysis of diet-disease associations showed that many of the 37 foods were associated with either one or even concurrently with several chronic diseases outcomes, including associations with MI, stroke, T2DM and smoking-related cancer but rarely with smoking-unrelated cancer. Inverse associations with disease emerged for whole-grain bread, breakfast cereals, vegetarian dishes, fresh fruits, compote, fruit juice, plant oils, vegetables (despite positive associations with stroke), wine (adjusted for alcohol intake), cheese, milk and milk products, butter, nuts, pasta/rice, pizza, sweet bread spreads, cakes and desserts. Positive associations with disease were observed for meat (particularly red and processed

meat), eggs, refined-grain bread, fried potatoes, margarine, cooked potatoes, stew, sauce, and confectionery.

Interestingly, the observed associations between foods and disease risks in the EPIC-Heidelberg and similar associations in many other epidemiologic studies seem to be explainable by a few general and major disease risk factors, mainly the level of adiposity and smoking. It is likely that also the residual diet-disease associations that typically remain in epidemiologic studies after adjustment for confounding variables are leftovers of confounder effects mainly due to incomplete adjustments for visceral/ectopic adiposity using anthropometric adiposity indices and for present and past smoking habits.

Indeed, the foods inversely and positively associated with disease risks appear to be interrelated components of more comprehensive behavioural patterns in subjects which might contribute to differences in body fat levels between subjects and might thus contribute to differences in the incidence of obesity-related diseases between subjects. Thus, in conclusion, the best dietary advice for the prevention of cardiovascular diseases, type 2 diabetes mellitus and obesity-related cancer might be a general change of the dietary intake habits of subjects with higher disease incidence to those of subjects with lower disease incidence for the prevention of overweight and obesity.