Clustering of dietary variables and other lifestyle factors (Dutch Nutritional Surveillance System)

Karin F A M Hulshof, M Wedel, M R H Löwik, F J Kok, C Kistemaker, R J J Hermus, F ten Hoor, Th Ockhuizen

Abstract

Study Objective—The aim was to investigate whether dietary factors cluster in a favourable or unfavourable way and to characterise the groups identified by lifestyle and sociodemographic variables.

Design and setting—This cross sectional study was based on data of the 1987-1988 Dutch national food consumption survey (DNFCS), obtained from a panel by a stratified probability sample of the non-institutionalised Dutch population.

Participants—3781 adults (1802 males and 179 females) of the DNFCS, aged 19 to 85 years, were studied.

Measurements and main results—To estimate dietary intake two day food records were used. Lifestyle factors were collected by structured questionnaire and sociodemographic variables were available from panel information. Cluster analysis was used to classify subjects into groups based on similarities in dietary variables. Subsequently, these groups were characterised by sociodemographic and lifestyle factors as well as by the consumption of food groups. Eight clusters were found. In comparison with the guidelines, the dietary quality in four clusters was poor. The cluster with the poorest dietary intake (high intake of fat, cholesterol, and alcohol; low intake of dietary fibre) showed on average a high consumption of animal products (except milk), fats and oils, snacks, and alcoholic beverages, and a low consumption of fruit, potatoes, vegetables, and sugar rich products. Smoking, body mass index, dietary regimen on own initiative, hours of sleep, gender, age, socioeconomic status, and day of the week were found to discriminate among the clusters.

Conclusions—Cluster analysis resulted in substantial differences in mean nutrient intake and seems useful for dietary risk group identification. Undesirable lifestyle habits were interrelated in some clusters, but an exclusive lifestyle for health risk has not been found.

J Epidemiol Community Health 1992; 46: 417-424

Lifestyle patterns contribute to morbidity and mortality in affluent societies. Although lifestyle variables, such as smoking, independently affect health status, interdependences among these factors are frequently observed. A combination of lifestyle practices may introduce a health risk that is greater than would be expected from the sum of

the individual factors.¹⁻⁴ Among adult males a relatively high consumption of coffee and alcohol, smoking, and few hours of sleep per night are all positively associated with body mass index (BMI).⁵ Prevalences of cardiovascular and other health related risk factors are often highest in lower socioeconomic strata in affluent societies.⁶⁻⁹ In The Netherlands, sociodemographic factors are found to be associated with diet and other health related variables (smoking, obesity, physical inactivity).¹⁰⁻¹² Combinations of cardiovascular risk factors are more frequently observed among elderly people with a lower socioeconomic status, whereas the associations between socioeconomic status and the separate risk factors are less consistent.¹³ Furthermore, differences in mean food consumption among Dutch population groups are relatively small. For instance, adults have a mean fat intake accounting for about 40% of total energy intake (40 en%), whereas the difference between high and low socioeconomic status groups is about 1.5 en%.¹⁴

Classifying people according to their consumption pattern by simultaneously considering the total diet may provide larger differences among groups based on nutrient profiles. The purpose of our study was thus to investigate whether dietary factors cluster in a favourable or unfavourable way and whether the groups identified could be characterised by lifestyle and sociodemographic variables. Subsequently, the consumption of food (groups) was studied to provide insight into the origin of the differences among clusters.

Methods

STUDY POPULATION AND DATA COLLECTION

In 1987-1988, the first Dutch national food consumption survey (DNFCS) was carried out. The data were obtained from a probability sample of the non-institutionalised Dutch population aged 1 to 85 years (n = 5898, response 81%). The methods regarding sampling procedures and dietary data collection are described in detail elsewhere.¹⁵ Briefly, to assess dietary intake two day estimated records were used. In each household the person principally responsible for domestic affairs was the most important informant and was visited on two occasions by a specially trained dietitian. During the first visit the household diary was explained. In this diary all the food supplied by the main housekeeper to the members of the household was recorded. Precise description of methods of cooking, recipes, and ingredients was requested. The housekeeper also noted for each meal the persons attending and the type and quantities of food items served to them, as well as the amounts of leftovers and food given

Department of
Human Nutrition,
TNO Toxicology and
Nutrition Institute,
PO Box 360,
3700 AJ Zeist,
The Netherlands
K F A M Hulshof
M Wedel
M R H Löwik
F J Kok
C Kistemaker
R J J Hermus
Th Ockhuizen
Department of
Human Biology, State
University Limburg,
Maastricht, The
Netherlands
F ten Hoor

Correspondence to:
Ms Hulshof

Accepted for publication
November 1991

to pets. In addition, a diary was kept by each person to record food eaten out. During the second visit the interviewers checked the diary for completeness. Common household measures and food regularly used were weighed. In the present study, people under 19 years were excluded because food consumption patterns for children and adolescents differ from those for adults. Pregnant women ($n = 52$) and subjects with a diet prescribed by a physician ($n = 353$) were excluded because these conditions are likely to affect food selection and hence nutrient intake. The population remaining for statistical analyses included 3781 adults (1802 men and 1979 women).

For each individual the average intake of foods and nutrients over two (consecutive) days was calculated. Nutritional supplements were not included in the calculations of nutrient intakes. Intake of protein, fat, carbohydrate, and alcohol was calculated as percentage of total energy intake (en%) and the intake of other nutrients was expressed per MJ. In total, 883 different food items (codes corresponding with the extended computerised version of the Dutch food composition table¹⁶) were used by the respondents. The costs for buying these products were estimated using data obtained from the Netherlands Central Bureau of Statistics.^{17, 18} In case of missing data, information on costs was obtained from large supermarkets and from a marketing research institute. Food items were combined into 26 food groups classified on the basis of similarity in nutrient composition and/or origin.

In addition to information on food consumption, data were collected on the respondent's body weight and standing height (both self reported values), use of nutritional supplements, special dietary practices, smoking habits, meal pattern, and hours of sleep on the two consecutive days, by means of structured questionnaires. Information about education, occupation, socioeconomic status, and demographic variables was obtained from the marketing research institute which carried out the field work (AGB-Attwood).

STATISTICS

K-means cluster analysis¹⁹ was used to classify individuals into a limited number of groups on the basis of similarity in nutrient intake. The criteria (dimensions) chosen to classify subjects were the consumption of fat (en%), P:S ratio (polyunsaturated fatty acids/saturated fatty acids), dietary fibre (g/MJ), cholesterol (mg/MJ), monosaccharides and disaccharides (en%), and alcohol (g). These variables are considered to have a potential impact on the health of the population in general, and for all these variables dietary guidelines have been formulated.²⁰ To restrict the influence of alcohol on the other variables (all related to energy) used in the clustering, the

calculation of energy was exclusive of the energy derived from alcohol.

The K-means cluster procedure of the BMDP statistical package (program KM)²¹ was used to construct the clusters. Because of differences in scale, the variables were standardised before clustering. The K-means cluster procedure requires that the number of clusters is specified before the analysis. As the actual number of clusters present in the data was unknown, the selection of the best number of clusters was based on a so called scree plot, in which the variance that remained within the clusters is portrayed against the number of clusters (varying from two to 10 clusters). When the plot levels off, no additional reduction of the within cluster variance is achieved. The number of clusters corresponding with this point was selected as the most appropriate fit to the data.

To obtain an impression of the nutritional quality of the diets of the identified clusters a scoring system based on the Dutch Dietary Guidelines²⁰ was developed. For each dietary variable a score of 1 was attached to an intake most deviant from the recommendation, whereas a score of 5 was given to an intake level considered to be most desirable (table I). Thereafter, a weighted sum of the scores for the different criteria was calculated for each cluster. This weighting was introduced since differences in potential impact on health exist. Actually, the highest weight (10) was given to fat; P:S ratio and dietary fibre were assigned a weight of 3, cholesterol a weight of 2, and monosaccharides and disaccharides as well as alcohol a weight of 1. In this way a sum score was obtained ranging potentially from 20 (indicating a diet of poor quality) to 100 (indicating a prudent diet). Although the weighing system developed is rather subjective, the system does not affect the results obtained and is used only as an aid in interpreting the results. A further motivation of the system and its robustness to changes in the subjective weights is provided in the discussion section.

The clusters found were characterised by sociodemographic and health related lifestyle factors,^{3-5, 22-25} as well as by the intake of food groups. Available variables were: smoking, obesity, breakfast habits, special dietary habits, use of nutritional supplements, alcohol consumption, coffee consumption, and hours of sleep. A (stepwise) discriminant analysis was performed to study whether these variables differed for the various clusters. The type of day (two week days, two weekend days, or one week and one weekend day) was incorporated in this analysis, because the distribution of the food record days may differ among clusters and could therefore influence average nutrient intake for the clusters. A second analysis was performed to find combinations of food group consumption that best discriminated among the clusters. The computed standardised discriminant function coefficients were used to indicate the relative importance of the variables for the discrimination among the clusters.

One of the drawbacks of a cluster analysis is the possible dependency of the results on the sample selected and the methodology used for dietary assessment. Other clusters may be found when another sample is used. Therefore we established

Table I The scoring system used to evaluate the nutritional quality of the diet in each cluster

Cluster variables	Score					Weight
	1	2	3	4	5	
Fat (en%) ^a	≥ 45	40–< 45	35–< 40	30–< 35	< 30	10
P:S ratio	< 0.35	0.35–< 0.40	0.40–< 0.45	0.45–< 0.50	≥ 0.50	3
Dietary fibre (g/MJ)	≤ 1.8	1.9–< 2.3	2.3–< 2.6	2.6–< 3.0	≥ 3.0	3
Cholesterol (mg/MJ)	≥ 44	38–< 44	33–< 38	26–< 33	< 26	2
Mono/disaccharides (en%)	≥ 30	26–< 30	22–< 26	18–< 22	< 18	1
Alcohol (g/day)	> 40	31–≤ 40	21–≤ 30	11–≤ 20	< 10	1

^a en% = % of total energy intake

Table II Daily intake of dietary variables^a and indication of nutritional adequacy according to eight clusters. Values are means (SD)

Dietary variable	1 (n = 461)	2 (n = 517)	3 (n = 425)	4 (n = 550)	5 (n = 515)	6 (n = 557)	7 (n = 490)	8 (n = 268)
Fat (en%) ^b	48.9 (5.3)	48.4 (4.9)	44.8 (5.2)	44.4 (4.5)	42.5 (4.1)	39.1 (4.1)	35.9 (5.1)	32.5 (5.6)
P:S ratio	0.37 (0.13)	0.44 (0.16)	0.77 (0.21)	0.34 (0.13)	0.36 (0.13)	0.42 (0.14)	0.36 (0.14)	0.41 (0.16)
Dietary fibre (g/MJ)	2.6 (0.7)	1.8 (0.4)	2.7 (0.7)	3.0 (0.6)	2.0 (0.4)	2.3 (0.6)	3.7 (1.0)	2.4 (0.8)
Cholesterol (mg/MJ)	40 (15)	44 (14)	26 (10)	38 (12)	36 (11)	28 (9)	31 (11)	23 (8)
Mono/disaccharides (en%)	14.6 (4.0)	16.4 (4.4)	19.6 (4.8)	17.4 (4.0)	24.3 (3.8)	24.6 (4.1)	23.4 (4.7)	33.8 (4.9)
Alcohol (g)	0.3 (1)	35 (23)	7 (13)	25 (18)	0.6 (2)	31 (22)	2 (5)	7 (15)
Sum score ^c	42	31	64	50	46	61	67	74

^a Calculated without the energy derived from alcohol.

^b en% = percentage of total energy intake.

^c Sum score has a minimum value of 20 points and a maximum of 100 points (see table I).

the stability of the cluster solution across independent samples. A clustering procedure identical to the procedure described above was performed on an independent sample of 395 subjects aged 65–80 years, for whom the habitual intake of the selected nutrients was estimated by means of a dietary history with cross check.²⁶ The average values of the criteria found in the clusters identified in the present study were used as initial cluster centres for this analysis. The correspondence of the cluster solutions was calculated for each cluster using correlation coefficients, after standardising the variables.

Results

GENERAL DESCRIPTION OF THE POPULATION

For both men and women the mean age was approximately 42 years. Almost half of the population belonged to the stratum with a low

socioeconomic status and 10% to the highest stratum. About 20% of the men and 13% of the women had a higher professional or university education. Sixteen per cent of the women and 6% of the men were living alone, 15% *v* 9% used nutritional supplements, 9% *v* 3% followed a dietary rule (eg, vegetarianism) or dietary regimen on their own initiative. Forty seven per cent of the men and 38% of the women were smokers. About 66% of the men and 44% of the women reported having consumed alcohol during the two day record period. In general, men were taller and heavier than women and had a higher body mass index (24.4 *v* 23.7 kg/m²).

CLUSTER ANALYSIS

Eight clusters were identified as the most adequate representation of the structure of dietary intake combinations in the data. Table II lists the mean values of the six criteria in the eight clusters, ranked by fat intake. In five clusters the mean contribution of fat to energy intake exceeded 40%. In clusters 1 and 2, representing 25% of the sample, the mean fat intake was close to 50 en%. In two clusters the high fat intake was accompanied by a high alcohol consumption and a low intake of monosaccharides and disaccharides (clusters 2 and 4). The lowest mean fat intake (32.5 en%), combined with the highest monosaccharide and disaccharide consumption (33.8 en%), was observed in cluster 8. Only one cluster (cluster 3) had a mean P:S ratio > 0.50; this value was accompanied by a relatively high fat intake. In comparison with the guidelines, the dietary quality of clusters 1, 2, 4, and 5 was moderate or poor, as indicated by the sum scores of 50 or lower. The average dietary characteristics among members of clusters 3, 6, 7, and 8 were more in accord with the recommendations, sum scores being 60 or higher. The highest energy intake was observed in clusters 2 and 6 (11.6 and 11.3 MJ/day, respectively), whereas clusters 1 and 7 had the lowest intake (9.0 and 8.2 MJ/day, respectively). Subjects in cluster 7 used diets with the highest mean mineral and vitamin density, whereas in cluster 2 the lowest density was observed (results not shown).

DISCRIMINANT ANALYSES

Variables with standardised discriminant function coefficients greater than 0.35 (or less than -0.35) for at least one of the canonical variables are presented in tables III and IV. For each of the analyses, three canonical variables had a clear interpretation and explained about 90% of the variance in the predictor variables. The standardised discriminant function coefficients in the first analysis suggest that gender, smoking, and weekdays were most predictive for the first canonical

Table III Standardised discriminant function coefficients and cluster means for canonical variables related to sociodemographic and lifestyle characteristics

Characteristics	Canonical variable ^a		
	1	2	3
Age (years)	0.02	0.25	0.56
BMI (kg/m ²)	0.08	-0.49	0.51
Gender (female)	0.61	-0.56	0.01
SES low	0.11	-0.42	-0.28
Smoking	-0.42	-0.31	0.11
Sleeping time (< 7 h)	-0.27	-0.11	0.43
Weekday	0.79	0.25	0.79
One weekday and one weekend day	0.35	0.04	0.33
Dietary rule/regimen on own initiative	0.18	0.50	0.07
Cluster means ^b			
Cluster 1	0.42	-0.32	0.11
Cluster 2	-0.55	-0.10	0.15
Cluster 3	0.11	0.15	-0.11
Cluster 4	-0.06	0.14	0.30
Cluster 5	0.12	-0.23	-0.17
Cluster 6	-0.51	0.09	-0.17
Cluster 7	0.59	0.23	0.01
Cluster 8	0.05	0.06	-0.30

^a The canonical variables constitute those linear combinations of the characteristics that best discriminate between the clusters.

^b Means of canonical variables for each cluster; only the relative magnitude of the means among the clusters can be interpreted.

BMI = body mass index; SES = socioeconomic status.

Table IV Standardised discriminant function coefficients and cluster means for canonical variables related to the intake of food groups

Food groups	Canonical variable ^a		
	1	2	3
Fruits	0.36	0.11	0.26
Bread	0.14	0.27	0.44
Cheese	-0.37	-0.03	-0.07
Eggs	-0.39	-0.01	-0.20
Fats and oils	-0.51	-0.05	-0.44
Sugar (products)	0.72	-0.55	-0.18
Alcoholic beverages	-0.12	-0.73	0.67
Meat/poultry fat	-0.50	-0.07	-0.24
Soft drinks	0.46	-0.30	-0.21
Cluster means ^b			
Cluster 1	-0.80	0.95	-0.44
Cluster 2	-1.78	-0.79	-0.08
Cluster 3	0.10	0.49	-0.19
Cluster 4	-0.57	0.35	0.63
Cluster 5	0.41	-0.19	-1.00
Cluster 6	0.40	-1.07	0.57
Cluster 7	1.02	0.96	0.54
Cluster 8	2.32	-0.78	-0.35

^{a,b} See table III

Figure 1 Mean daily consumption of various food groups by cluster expressed as the deviation (%) from the overall consumption, the circle corresponding with the average consumption of the total population. HFLA=high fat/low alcohol cluster; MFLA=medium fat/low alcohol cluster; HFHA=high fat/high alcohol cluster.

variable, which explained 64% of the variance of the sociodemographic and lifestyle characteristics. The cluster means of the discriminant scores show that the first variable predominantly discriminated clusters 1 and 7 from clusters 2 and 6. Body mass index, gender, low socioeconomic status, and dietary regimen were strongly associated with the second canonical variable, which explained another 13% of the variance and discriminated clusters 1 and 5 from cluster 7. Age, body mass index, hours of sleep, and weekdays were most important for the third canonical variable, which explained 14% of the variance and discriminated between clusters 4 and 8. Region, urbanisation level, education, household size, missing breakfast during the survey, and use of supplements hardly discriminated among the clusters, all having low standardised discriminant function scores.

In the second stepwise discriminant analysis, the consumption of 26 food groups was used to explain cluster differences. The results indicated that 51% of the variance in the food groups was explained by the first canonical variable. Sugar rich products, fats and oils, high fat meat, soft drinks, eggs, cheese, and fruit were most predictive for this variable, which discriminated clusters 1 and 2 from clusters 7 and 8 (table IV). The second canonical variable, with sugar rich products and alcoholic beverages, explained 26% of the total variance and discriminated clusters 2, 6, and 8 from clusters 1 and 7. The third canonical variable with bread, fats and oils, and alcoholic beverages as most predictive variables, explained 14% of the variance and discriminated between cluster 5 and clusters 4, 6 and 7.

CLUSTER CHARACTERISTICS

A more detailed description of the characteristics of clusters 1, 2, and 7 will be presented, as these clusters constitute the most relevant differences in the food intake, sociodemographic and lifestyle variables (see discriminant analyses). Based on two of the cluster criteria the clusters are called high fat/low alcohol (cluster 1), moderate fat/low alcohol (cluster 7), and high fat/high alcohol (cluster 2).

Sociodemographic and lifestyle variables

Table V summarises the sociodemographic and lifestyle characteristics that differ significantly among the clusters (χ^2 test). The high fat/low alcohol cluster contained a higher proportion of women, more subjects with low socioeconomic status, more subjects with (severe) obesity, and an overrepresentation of weekdays. The moderate fat/low alcohol cluster comprised more participants with high socioeconomic status, more with a dietary regimen on their own initiative, more women, fewer smokers, and fewer subjects with less than 7 h sleep. This cluster also contained a higher proportion of subjects who had reported dietary intake during weekdays. The high fat/high alcohol cluster was made up of a higher proportion of men, more middle aged participants, more smokers, more subjects with less than 7 h sleep, and a higher proportion of subjects that reported dietary intake on one or two weekend days.

Table V Sociodemographic and lifestyle characteristics according to cluster

Cluster	HFLA ^a (n=461)	MFLA (n=490)	HFHA (n=517)	All (n=3781)
Proportion of women (%)	69	67	39	52
Age (years)	42.8	42.6	41.9	41.6
Age group (%): 19–24 years	10	11	8	11
25–49 years	58	56	64	60
50–64 years	22	22	19	19
≥65 years	10	11	9	10
SES (%): low	57	41	45	47
high	7	13	10	10
Education (%): low	49	35	40	40
high	8	20	16	16
Body mass index (%): obese (25.0–29.9 kg/m ²)	28	31	26	27
severely obese (≥30 kg/m ²)	10	6	5	5
Smoking (%)	39	24	54	42
Dietary regimen (%)	5	13	4	6
Type of surveyed days (%):				
2 weekdays	69	71	41	59
2 weekend days	8	7	23	14
1 weekday and 1 weekend day	23	21	36	27
Sleeping time (%):				
<7 h	9	4	16	10
7–8 h	43	47	36	43
≥9 h	48	48	48	47
Use of breakfast (%):				
both days	85	90	73	82
1 day	6	5	13	9
0 days	9	5	13	9
Nutritional supplement usage (%)	12	16	8	12
Living alone (%)	10	14	8	11

^a HFLA = high fat/low alcohol; MFLA = moderate fat/low alcohol; HFHA = high fat/high alcohol; SES = socioeconomic status.

Food consumption pattern

Figure 1 presents the deviation (in per cent) from the overall mean consumption of relevant food groups per cluster. The high fat/low alcohol cluster consisted of subjects with a somewhat higher mean consumption of potatoes and animal products (especially high fat meat), whereas the intake of products mainly consumed between the meals (snacks, sugar rich products, soft drinks, and alcoholic beverages), mixed dishes, and milk (products) was low. Members of the moderate fat/low alcohol cluster consumed more potatoes, vegetables, and fruits, and more (low fat) milk (products) and tea than the average population, whereas the consumption of all other food items, bread and pastries excepted, was (much) lower. The high fat/high alcohol cluster had a high intake of animal products (except milk), fats and oils, savoury snacks, mixed dishes, and alcohol (almost half of the cluster members had used more than six glasses of alcoholic beverages during the two days). This intake pattern was accompanied with a (very) low consumption of fruit, potatoes, vegetables, sugar rich products, and tea.

Costs of food

Figure 2 shows the average daily costs according to cluster and the contribution of different food groups to these costs. Including alcohol, the mean

Figure 2 Average daily costs of diets; contribution (%) of food groups to costs.

daily costs for foods and drinks ranged from DfI (Dutch guilders) 7.06 (high fat/low alcohol) to DfI 10.28 (high fat/high alcohol). In all clusters the group "meat and poultry" was the most important contributor to dietary costs (22–31%). In the three clusters with a high alcohol consumption, alcoholic beverages were responsible for 19–22% of the costs. Other important contributors were vegetables, milk (products), cheese, and non-alcoholic beverages such as soft drinks and coffee. When the costs of alcoholic beverages were excluded from the calculations the food patterns of subjects in the high fat/high alcohol cluster remained the most expensive. However, when the daily costs were related to energy intake the picture changed somewhat. The average costs per MJ in the moderate fat/low alcohol cluster approached those for the high fat/high alcohol cluster.

STABILITY OF CLUSTER SOLUTIONS

The results of the validation study revealed that the overall agreement between cluster profiles was reasonably good ($r = 0.62$ on average). The most marked differences were observed for cholesterol intake and P:S ratio, especially in clusters 1 and 3. In these two clusters the correspondence between the average cluster criteria based on the DNFCs and the validation results was low (both $r = -0.21$). In the other clusters the correlation coefficients varied from 0.71 (cluster 4) to 0.96 (cluster 7). When the cluster means were calculated for subjects aged 65 years or over in the DNFCs the agreement between the cluster solutions hardly differed.

Discussion

In this study we have used cluster analysis in an attempt to classify individuals on the basis of their nutrient intake into rather homogeneous groups. As compared to the more traditional approach of a priori selection of groups (eg, socioeconomic groups), followed by analysis of variance, cluster analysis examines more dynamically patterns of intake of (selected) nutrients by grouping subjects with comparable combinations of intake levels. Cluster analysis or factor analysis based on the individual consumption of foodstuffs is used to identify eating patterns.^{27–32} However it is not the presence or absence of separate food items but the appropriate selection (proper amounts and combinations) of foods that is important to health. Therefore, we grouped respondents on the basis of dietary factors for which guidelines have been formulated.

Smoking, body mass index, dietary regimen, hours of sleep, gender, age, socioeconomic status, and type of surveyed day were found to discriminate among the clusters. A priori segmentation by socioeconomic status, age, and gender separately did show some differences in dietary intake, but they were small.¹⁴ In this study differences in mean levels found among clusters were much larger and combinations of (un)desirable factors were observed. For example, the high fat/high alcohol cluster contained more smokers and more people who slept less than 7 h per night as compared to the overall population. Thus cluster analysis based on dietary characteristics is a promising method for risk group

identification. However, on the basis of a limited number of lifestyle variables and traditional sociodemographic factors, we could not find an exclusive lifestyle for specific groups of the population. Further study is warranted to determine whether there are indices of lifestyle and or sociodemographic characteristics which are more sensitive and therefore could provide a better segmentation of the population.

As the possible instability of the results could be one of the limitations of a cluster analysis, we studied the validity of the cluster solution in another sample. This sample was made up of subjects aged 65–80 years and figures for the usual dietary intake were available. The overall correspondence between average values of the clustering criteria within clusters was satisfactory. Future studies should provide further evidence of the stability of the results of the present study.

As in other studies, gender as well as day of the week influence food consumption.^{30 31 33} Instead of performing cluster analyses separately for males and females and for weekdays and weekend days we used both factors in discriminant analysis. In this way we avoided an increase in the number of clusters and a decrease in the size of these clusters which would have led to complex results that are less stable and more difficult to interpret. Gender and type of day were not always associated with differences in (un)desirable lifestyle variables found among clusters. For instance, the high fat/low alcohol and the moderate fat/low alcohol clusters differed substantially in intake levels but not in gender or percentage of weekdays. A high alcohol consumption in the different clusters was associated with weekend days as well as with weekdays (the latter results are not presented). The differences among the clusters regarding the day of the week represented may have important implications for the interpretation of the results. The clusters found may partly be the result of somewhat "unhealthier" food habits during the weekends. Since the day of the week was hardly associated with the second canonical variable that discriminated between high fat/low alcohol and moderate fat/low alcohol, clusters were not solely the result of type of day. Moreover, the results obtained from the dietary intake figures in the second sample reflecting habitual food consumption were in line with the results based on two day records.

The nutritional health of the average diets of the various clusters was evaluated with the aid of a scoring system. In this system criteria of a prudent diet were used as a standard.^{20 34} In The Netherlands, a high fat intake is currently considered to be the main nutritional problem and discouragement of a high fat intake is generally accepted to be of prime importance. Therefore the fat score is given the highest weight (10 points). An excessive alcohol consumption constitutes a serious health risk and is discouraged. Therefore alcohol consumption could receive a relatively high weight in the system. However, the sample we used was not likely to include individuals with excessive alcohol intake.¹⁵ Although some health risks of moderate alcohol consumption have been identified, beneficial effects have also been reported³⁵ and alcohol was thus given a weight of 1 point. Of course, the introduction of weights is

somewhat subjective. However, when fat was given a weight of 10 points and all the other criteria were given 3 points, the ranking of the clusters based on their sum scores was identical to that presented in table III. This was also the case when we used the intake of saturated fatty acids (en%) instead of the P:S ratio. In a recent update of the guidelines for a healthy diet in 1991, the Netherlands Nutrition Council recommends that saturated fatty acids should not exceed 10 en%, whereas no special guidelines for the P:S ratio are given any more.³⁶ In all clusters the mean contribution of saturated fatty acids to energy intake was much higher than 10% [range 13.5% (cluster 8) to 20.6% (cluster 1)]. These results suggest that the overall assessment of the cluster solutions with respect to risks for chronic diseases seems quite stable.

On the basis of total sum scores, more than 50% of the examined population had diets that departed substantially from the guidelines for a healthy diet. As pregnant women and subjects with a diet prescribed by a physician were excluded from cluster analysis the results cannot be generalised to the whole adult population. In comparing the results of our study with those of others, it should be realised that the methods used to collect and classify data vary among studies. Moreover, differences among dietary patterns between countries could also introduce cluster differences. Despite these limitations, results of studies abroad support several of our observations with regard to eating patterns and their characteristics. For example, another cluster analysis based on the nutrient density of diets consumed by US women aged 19–51 years classified the population into nine groups. More than half of this US sample used diets of poor nutritional quality.³⁷ Low energy diets of poor nutritional quality were associated with women in households with lower income levels and with female heads of households with lower level of education.³⁷ Indicators of these characteristics were also overrepresented in our high fat/low alcohol cluster. Aktin *et al*³⁰ reported that with the aid of cluster analysis based on food intake the food patterns of elderly people can be well categorised as light eaters, heavy eaters, or consumers of large amounts of alcoholic beverages, salty snack products, animal fat products, and so on, which agrees with some of the food group characteristics in our clusters, for instance in the high fat/high alcohol cluster.

Our results indicate that the moderate fat/low alcohol cluster can be identified as the cluster with the most prudent lifestyle, and this may partially be motivated by health considerations among its members. In this cluster subjects with high socioeconomic status were overrepresented. Other studies have also shown that a more pronounced health conscious behaviour is often associated with a higher social status,^{38–40} resulting in a more favourable biochemical profile of cardiovascular risk factors.³¹ Theoretically, subjects with a higher socioeconomic status could spend more money on food. Mooney⁴¹ reported that healthy food choices appear to be more expensive. In our study the costs of the prudent diet of those in the moderate fat/low alcohol cluster, associated with high socioeconomic status, did not differ substan-

tially from the costs of a poorer diet in the high fat/low alcohol cluster, in which more people were of low social status. In the clusters with a high alcohol consumption the diets were on average the most expensive (and remained so if the costs of alcoholic beverages were not taken into account), but no association with socioeconomic status was found. These results have to be interpreted with caution, because both clusters with a high alcohol intake also had the highest energy intake ("largest eaters"). Comparing the costs per MJ revealed that the costs of the diets are influenced not only by the quantity but also by quality.

In conclusion, our results show that cluster analysis can identify groups with noteworthy differences in mean nutrient intake and this method can be considered as a useful tool for dietary risk group identification. Undesirable lifestyle variables are interrelated in some clusters, but an exclusive lifestyle for health risk has not been found. At this moment our study does not provide a good basis for delivering diversified recommendations for specific segments of the population. Future studies should try to find more sensitive indices that discriminate nutritional risks with a higher degree of specificity which may result in targeting of nutrition messages.

This study was carried out within the framework of the Programme Committee Socioeconomic Health Differences and was supported financially by the Ministry of Welfare, Public Health and Cultural Affairs (grant 8963). We are very grateful to Evelien Aarnink for assistance in data processing and computer analysis. The comments of Dirk G van der Heij are acknowledged with appreciation.

- 1 Breslow L, Enstrom JE. Persistence of health habits and their relationship to mortality. *Prev Med* 1980; 9: 469–83.
- 2 Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 1981; 6: 1191–308.
- 3 Wingard DL, Berkman LF, Brand RJ. A multivariate analysis of health-related practices. *Am J Epidemiol* 1982; 116: 765–75.
- 4 Belloc NB, Breslow L. Relationship of health status and health practices. *Prev Med* 1972; 1: 409–21.
- 5 Baecke JAH, Burema J, Frijters JER, Hautvast JGAJ, Van der Wiel-Wetzels WAM. Obesity in young adults: II. Daily life-style and body mass index. *Int J Obesity* 1983; 7: 13–24.
- 6 Baecke JAH, Burema J, Frijters JER, Hautvast JGAJ, Van der Wiel-Wetzels WAM. Obesity in young Dutch adults: I. Socio-demographic variables and body mass index. *Int J Obesity* 1983; 7: 1–12.
- 7 Holme I, Helgeland A, Hjermann I, Lund-Larsen PG, Leren P. Coronary risk factors and socio-economic status —The Oslo study. *Lancet* 1976; ii: 1396–9.
- 8 Noppa H, Bengtsson C. Obesity in relation to socio-economic status. *J Epidemiol Community Health* 1980; 34: 139–42.
- 9 Duijkers TJ, Kromhout D, Spruit IP, Doornbos G. Intermediating factors in the relation between socioeconomic status and 25-year mortality (The Zutphen Study). *Int J Epidemiol* 1989; 18: 658–62.
- 10 Kok FJ, Matroos AW, Van den Ban AW, Hautvast JGAJ. Characteristics of individuals with multiple behavioural risk factors for coronary heart disease in The Netherlands. *Am J Public Health* 1982; 72: 986–91.
- 11 Hulshof KFAM, Wedel M, Ockhuizen Th. Dietary intake, lifestyle and anthropometry of 18-year old men in The Netherlands. *Nutr Rep Int* 1988; 37: 755–65.
- 12 Löwik MRH, Westenbrink S, Wedel M, Schrijver J. De invloed van sociaal economische factoren op de voeding en voedingstoestand van ouderen. *Voeding* 1989; 50: 38–42.
- 13 Löwik MRH, Borsboom AGB, Bun CJE, Kok FJ, Schouten EG, Odink J. Clustering cardiovasculaire risico-indicatoren bij ouderen. *Hart Bull* 1991; 22: 95–9.
- 14 Hulshof KFAM, Löwik MRH, Kok FJ, Wedel M, Brants HAM, Hermus RJJ, Ten Hoor F. Diet and other life-style factors in different socio-economic groups (Dutch Nutrition Surveillance System). *Eur J Clin Nutr* 1991; 45: 441–50.
- 15 Hulshof KFAM, Van Staveren WA. The Dutch National Food Consumption Survey: design, methods and first results. *Food Policy* 1991; 16: 257–60.

- 16 NEVO-Tabel. *Nederlands Voedingsstoffenbestand 1986-1987*. The Hague: Voorlichtingsbureau voor de Voeding, 1987.
- 17 Centraal Bureau voor de Statistiek. *Bijvoegsel Maandstatistieken van de prijzen*. Voorburg: CBS, 1987: 12.
- 18 Centraal Bureau voor de Statistiek. *Bijvoegsel Maandstatistieken van de prijzen*. Voorburg: CBS, 1988: 13.
- 19 Hartigan JA. *Clustering algorithms*. New York; 1975.
- 20 Netherlands Nutrition Council. *Guidelines for a healthy diet*. The Hague: Voedingsraad, 1986.
- 21 Dixon WJ, Brown MB, Engelsman L, et al. *BMDP statistical software*. Berkeley: University of California Press, 1990.
- 22 Axelson ML. The impact of culture on food-related behaviour. *Annu Rev Nutr* 1986; 6: 345-63.
- 23 Lubben JE, Weiler PG, Chi J. Health practices of the elderly poor. *Am J Public Health* 1989; 79: 731-4.
- 24 Shaw M, Jeffery RW, Hannan PJ, Onstod L. Relationship between socio-demographic and behaviour variables, and body mass index in a population with high-normal blood pressure: hypertension prevention trial. *Eur J Clin Nutr* 1989; 43: 583-96.
- 25 Solvoll K, Selmer R, Loken EB, Foss OP, Trygg K. Coffee, dietary habits, and serum cholesterol among men and women 35-49 years of age. *Am J Epidemiol* 1989; 129: 1277-88.
- 26 Löwik MRH, Westenbrink S, Hulshof KFAM, Kistemaker C, Hermus RJJ. Nutrition and aging: dietary intake of "apparently healthy" elderly. *J Am Coll Nutr* 1989; 8: 347-56.
- 27 Cronin FJ, Krebs-Smith SM, Wyse BW, Light L. Characterizing food usage by demographic variables. *J Am Diet Assoc* 1982; 81: 661-73.
- 28 Schwerin HS, Stanton JL, Riley AM, et al. Food eating patterns and health: a reexamination of the Ten-State and HANES I surveys. *Am J Clin Nutr* 1981; 34: 568-80.
- 29 Schwerin HS, Stanton JL, Smith JL, Riley AM, Brett BE. Food eating habits, and health: a further examination of the relationship between food eating patterns and nutritional health. *Am J Clin Nutr* 1982; 35: 1319-25.
- 30 Akin JS, Guilkey DK, Popkin BM, Fanelli MT. Cluster analysis of food consumption patterns of older Americans. *J Am Diet Assoc* 1986; 86: 616-24.
- 31 Nicklas TA, Webber LS, Thompson B, Berenson GS. A multivariate model for assessing eating patterns and their relationship to cardiovascular risk factors. *Am J Clin Nutr* 1989; 49: 1320-7.
- 32 Randall E, Marshall J, Graham S, Brasure J. Patterns in food use and their associations with nutrient intakes. *Am J Clin Nutr* 1990; 52: 739-45.
- 33 Beaton GH, Milner J, Corey P, et al. Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr* 1979; 32: 2546-59.
- 34 Netherlands Nutrition Council. *Richtlijnen goede voeding*. Voeding 1990; 47: 159-81.
- 35 Boffetta P, Garfinkel L. Alcohol drinking and mortality among men enrolled in an American Cancer Society Prospective Study. *Epidemiology* 1990; 1: 342-8.
- 36 Netherlands Nutrition Council. *Nader advies inzake de richtlijn m.b.t. de vetconsumptie uit het advies Richtlijnen goede voeding 1986*. The Hague, 1991.
- 37 Windham CT, Windham MP. Factors related to nutrient intake patterns of female participants in the USDA CSFII-1985 (abstract). *J Am Diet Assoc* 1990; 90: (suppl) A-94.
- 38 Anonymous. Use of vitamin and mineral supplements in the United States. *Nutr Rev* 1990; 48: 161-2.
- 39 Block G, Cox C, Madans J, Schreiber GB, Licitra L, Melia N. Vitamin supplement use, by demographic characteristics. *Am J Epidemiol* 1988; 127: 297-309.
- 40 Koplan JP, Annett JL, Layde PM, Rubin GL. Nutrient intake and supplementation in the United States (NHANES II). *Am J Public Health* 1986; 76: 287-9.
- 41 Mooney C. Cost and availability of healthy food choices in a London health district. *J Hum Nutr Diet* 1990; 3: 111-20.